

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 14:47:39 ; Search time 53.42 Seconds
(without alignments)
95.646 Million cell updates/sec

Title: us-09-833-017-2

Sequence: 1 MKRTLSIKNDKFKIKTDELE.....GSLSTFFRLNRSFTQALGK 46

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

A.GeneSeq_032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	25.6	173	18	AAW28307
2	56	24.7	939	22	AAU35204
3	55.5	24.4	553	8	AAW70176
4	55.5	24.4	553	15	AAW58858
5	55.5	24.4	553	15	AAW91141
6	55.5	24.4	553	17	AAW06828
7	55.5	24.4	553	18	AAW10691
8	55.5	24.4	553	20	AAW21983
9	55.5	24.4	553	21	AAW36039
10	55.5	24.4	553	21	AAW51231
11	55.5	24.4	553	21	AAW58183

12	55	24.2	583	22	AAW82132	S. epidermidis ope
13	54.5	24.0	508	21	AAW44399	Arabidopsis thalia
14	54.5	24.0	515	21	AAW44398	Arabidopsis thalia
15	54	23.8	96	14	AAW35757	Monellin protein.
16	54	23.8	97	14	AAW34182	Monellin protein.
17	54	23.8	2353	17	AAW93933	Haemophilus adhesi
18	54	23.8	2411	21	AAW23860	Haemophilus influe
19	53.5	23.6	553	12	AAW10065	Newcastle disease
20	53	23.3	96	21	AAW57201	Arabidopsis thalia
21	53	23.3	513	15	AAW54214	L. lactis branched
22	52	22.9	148	22	AAW28821	Novel human diago
23	52	22.9	365	22	AAW78620	Human protein S80
24	52	22.9	381	22	AAW79604	Human protein S80
25	52	22.9	381	22	AAW79605	Human protein S80
26	52	22.9	382	22	AAW78621	Human protein S80
27	52	22.9	516	21	AAW52007	M. jannaschii M08
28	52	22.9	516	21	AAW51636	M. jannaschii M08
29	51.5	22.7	177	22	AAW17731	Novel human diago
30	51.5	22.7	257	22	AAW98329	Escherichia coli p
31	51.5	22.7	306	21	AAW59918	E. coli proliferat
32	51.5	22.7	306	22	AAW59918	E. coli proliferat
33	51.5	22.7	591	22	ABG10345	Novel human diago
34	51.5	22.7	2737	22	ABG18288	Novel human diago
35	51.5	22.7	3076	22	ABG07038	Novel human diago
36	51	22.5	220	22	AAW37294	Staphylococcus aur
37	51	22.5	292	22	AAW74939	Human colon cancer
38	51	22.5	313	18	AAW21821	Mutant uracil DNA
39	51	22.5	313	18	AAW21822	Mutant uracil DNA
40	51	22.5	313	18	AAW21824	Mutant uracil DNA
41	51	22.5	313	18	AAW21825	Mutant uracil DNA
42	51	22.5	313	18	AAW21814	Uracil DNA glycosy
43	51	22.5	313	18	AAW21815	Cytosine DNA glyco
44	51	22.5	313	18	AAW21816	Thymine DNA glycos
45	51	22.5	313	18	AAW21819	Mutant uracil DNA

ALIGNMENTS

RESULT 1	AAW28307	standard: Protein: 173 AA.
ID	AAW28307	
XX	AAW28307	
AC	AAW28307	
XX		
DT	01-SEP-1998	(first entry)
XX		
DE	Staphylococcus aureus	protein of unknown function.
KW	Staphylococcus aureus	protein: ribozyme; antisense sequence; control;
KW	Staphylococcal gene; regulatory element; bacterial	gene expression;
KW	vaccine; Staphylococcal infection; food poisoning;	scalded skin syndrome;
KW	toxic shock syndrome.	
XX		
OS	Staphylococcus aureus.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 123	/note= "X is not defined in the specification"
FT	Misc-difference 132	/note= "X is not defined in the specification"
FT	Misc-difference 132	/note= "X is not defined in the specification"
XX		
PN	W09730070-A1.	
XX		
PD	21-AUG-1997.	
XX		
PF	19-FEB-1997;	97WO-US02318.
XX		
PR	20-FEB-1996;	96US-0011888.
XX		
PA	(SMK) SMITHKLINE BEECHAM CORP.	
XX		
PI	Black MT, Burnham MK, Hodgson JB, Knowles DJC, Nicholas RO;	

XX

PS Example: pages 11-16: 22pp: English

XX

XX		04-MAR-1997	(first entry)	
DT		Newcastle disease virus F gene product.		
XX	DE	Newcastle disease virus F gene product.		
XX		Turkey herpes virus; recombinant virus; vaccine; prophylaxis;		
KW		immunisation; avian virus; infectious bronchitis virus;		
KW		infectious bursal disease virus; Newcastle disease virus;		
KV		Marek's disease virus; infectious laryngotracheitis virus; IBV;		
KM		IBDV; NDV; MDV; ILV.		
XX	OS	Newcastle disease virus.		
XX	PN	MO605291-A1.		
XX	PD	22-FEB-1996.		
XX		09-AUG-1995; 95WO-US10245.		
PF		22-DEC-1994; 94US-0362240.		
PR		09-AUG-1994; 94US-0288065.		
XX		(SYTR) SYNTRO CORP.		
PA		Cochran MD, Junker DE, Singer PA, Wild MA;		
PI		WPI; 1996-139689/14.		
DR		N-PSDB; AAT18203.		
XX		Recombinant turkey herpes viruses contg. foreign DNA encoding a		
PT		cytokine - useful in vaccines to protect against Marek's disease		
PT		virus and other avian viruses.		
PS		Disclosure; Page 179-180; 249pp; English.		
XX		Recombinant turkey herpes virus (rTH) which comprise a foreign DNA		
CC		sequence encoding a cytokine inserted into a XhoI site within an		
CC		EcoRI #9 genomic fragment, where the cytokine can be expressed in		
CC		host cells infected with the virus can be used in vaccines to		
CC		protect turkeys against avian viruses. The recombinant viruses can		
CC		be used for immunising birds against infectious bronchitis virus		
CC		(IBV), infectious bursal disease virus (IBDV), Marek's disease virus		
CC		(MDV), infectious laryngotracheitis virus (ILV) and Newcastle disease		
CC		virus (NDV). They may also be used in multivalent vaccines to		
CC		protect against two or more of these avian viruses. This sequence		
CC		is the product of the F gene of the Newcastle disease virus and is		
CC		an antigen which can be used in the recombinant vaccines.		
CC				
XX	SQ	Sequence . 553 AA:		
	Query Match	24.4%; Score 55.5; DB 17; Length 553;		
	Best Local Similarity	33.3%; Pred. No.29;		
	Matches 16; Conservative	7; Mismatches 20; Indels 5; Gaps		
Oy	4 TLSTKDNF-----KEIKTDELRLITIGSGSLSTFFRLFNRKSFQTALGK 46	: : : : :		
Dd	433 tlrtsgefdvtqknlsiqdsqylvltfgndlstdelgnvnmsistnainlk 480	: : : : :		
RESULT 7				
AAAM10691				
ID	AAAM10691 standard; Protein; 553 AA.			
AC	AAAM10691;			
DX	05-MAY-1997 (first entry)			
XX	Newcastle disease virus fusion protein.			
DE	Fowlpox virus; FPV; recombinant virus; vector; vaccine;			
KW	immunisation; NDV; haemagglutinin; fusion protein; antigen;			
KW	poultry.			

XX Newcastle disease virus.
OS
XX
XX W09640880-A1.
FN
XX
XX 19-DEC-1996.
PD
XX
XX 04-JUN-1996; 96WO-US11187.
PF
XX 07-JUN-1995; 95US-0484730.
PR
XX
XX (SYTR) SYNTRO CORP.
PA
XX
XX Cochran MD, Junker DE, Singer PA;
PI
XX
XX WPI; 1997-087060/08.
DR
XX N-PSDB; AAT48510.
XX
PT New recombinant fowlpox virus - conty. a foreign DNA sequence
PI inserted into the fowlpox virus genome, used for the prodn. of
PT vaccines.
XX
XX
PS Disclosure: Page 109-110; 134pp; English.
XX
XX
XX Newcastle disease virus (NDV) haemagglutinin (HN) (AAM10690) and
CC fusion (F) protein (AAW10691) are expressed by novel recombinant
CC fowlpox virus (FPV). The genes (see also AAT48510) for HN and F can
CC be inserted into homology vector 443-88.8 (see also AAT48511) at the
CC unique SfiI site, yielding homology vector 502-26.22 (see also
CC AAT48502-05), which can be used to insert the NDV HN and F genes
CC into fowlpox virus. The recombinant virus is used to deliver the
CC vaccine antigens to poultry.
XX
SQ Sequence 553 AA;

```

Query Match      24.4%; Score 55.5; DB 18; Length 553;
Best Local Similarity 33.3%; Pred.No.29;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1

QY      4 TSLKNDK-----KETKDELEIIIGSGSLSTFFRLFNKSPQALGK 46
        ||| |:| | |::||| |:| | | |}
Db       433 FLIRSGFDVLYqknlsiqsgvILtgnldisteIqnvmsisnalnk 480

RESULT      8
AA21983
ID   AA21983 standard; Protein; 553 AA.
XX
AC   AA21983;
XX
DT   07-SEP-1999 (first entry)
XX
DE   Seq ID No: 14 of US5925358.
XX
Fowlpox virus; FPV; recombinant; vaccine; immunisation; chicken; NDV;
Newcastle disease virus; NDV; Fowlpox; Infectious laryngotracheitis.
XX
OS   Fowlpox virus.
XX
US5925358-A.
PN
XX
PD   20-JUL-1999.
XX
PF   07-JUN-1995;    95US-0484575.
XX
PR   07-JUN-1995;    95US-0484575.
PR   26-FEB-1993;    93US-0024156.
PR   28-FEB-1994;    94WO-US02252.
XX
PA   (SVTR ) SYNPRO CORP.
XX
PI   Cochran MD, Junker DE;

```


[illegible][illegible]

PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151080.
PR	31-AUG-1999;	99US-0151303.
PR	01-SEP-1999;	99US-0151438.
PR	07-SEP-1999;	99US-0151930.
PR	10-SEP-1999;	99US-0152363.
PR	13-SEP-1999;	99US-0153070.
PR	15-SEP-1999;	99US-0153758.
PR	16-SEP-1999;	99US-0154018.
PR	20-SEP-1999;	99US-0154039.
PR	22-SEP-1999;	99US-0154779.
PR	23-SEP-1999;	99US-0155139.
PR	24-SEP-1999;	99US-0155486.
PR	28-SEP-1999;	99US-0155659.
PR	29-SEP-1999;	99US-0156458.
PR	04-OCT-1999;	99US-0156596.
PR	05-OCT-1999;	99US-0157117.
PR	06-OCT-1999;	99US-0157753.
PR	07-OCT-1999;	99US-0157865.
PR	08-OCT-1999;	99US-0158029.
PR	12-OCT-1999;	99US-0158232.
PR	13-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	14-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	29-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

RESULT	14
AA644398	
ID	AA644398 standard; Protein: 515 AA.
XX	
AC	AA644398;
XX	
D7	18-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 55611.
XX	
KW	Protein identification; signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EPI033405-A2.
XX	
PD	06-SEP-2000.
XX	
FE	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	23-APR-1999; 99US-0130891.
PR	28-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	30-APR-1999; 99US-0132407.
PR	04-MAY-1999; 99US-0132484.
PR	05-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	06-MAY-1999; 99US-0132487.
PR	07-MAY-1999; 99US-0132863.
PR	11-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134221.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	21-MAY-1999; 99US-0135353.
PR	24-MAY-1999; 99US-0135628.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	14-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145099.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148179.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154479.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 24.0%; Score 54.5; DB 21; Length 515;
Best Local Similarity 39.6%; Pred. No. 36;
Matches 21; Conservative 8; Mismatches 11; Indels 13; Gaps 5;

QY 1 MKKTLISKND-FKE-----IKTDELLEIII-GSGSGSL--TF--FRLFNRSP 40
| | : | | | | : | | | | : | | | | : | | : |
Db 8 mekrstlkndsfvgeynpvtetgslstivlgaagdlakktfipalfnlngtf 60

RESULT 15

AAK35757
ID AAK35757 standard; Protein; 96 AA.

XX AAK35757;

AC AAK35757;
XX 29-JUL-1993 (first entry)
DT

```
XX Monellin protein.
DE
XX Sweet; heat; resistance; protease.
KW
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Misc-difference 51
FT /label= GLY, SER, GLU
XX
XX JP05070494-A.
PN
XX
XX 23-MAR-1993.
PD
XX
XX 30-MAY-1991; 91JP-0155713.
PF
XX
XX 25-JUL-1990; 90JP-0196983.
PR
XX
XX (KIRI ) KIRIN BREWERY KK.
PA
XX
XX WPI; 1993-131302/16.
DR
XX
XX Single-stranded monellin protein - has high heat resistance in
PT cow pH range and high protease resistance
XX
XX Claim 1; Page 2; 29pp; Japanese.
PS
XX
XX Monellin, is a sweet protein, having high thermal resistance at low
CC pH's and also a high protease resistance.
XX
XX Sequence 96 AA;
SQ

Query Match 23.8%; Score 54; DB 14; Length 96;
Best local Similarity 41.4%; Pred. No. 6;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

OY 1 MKKTLSLKNDPKFKIKDELEIITGGSGSL 29
   ||||: :| :||| | : : | |
Db 42 mkktiyeenxfreikgyeyqlvyvasdki 70
```

Search completed: July 30, 2002, 15:14:11
Job time: 1592 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 30, 2002, 15:06:04 ; Search time 23.35 seconds
(without alignments)
48.119 Million cell updates/sec

Title: US-09-833-017-2

Sequence: 1 MKKTSLKNDKFEKITDELE.....GSLSTFRLNRSFTQALGK 46

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.5	24.4	553	2 US-08-663-566A-13	Sequence 13, Appl
2	55.5	24.4	553	2 US-08-484-575A-14	Sequence 14, Appl
3	55.5	24.4	553	2 US-08-023-610-13	Sequence 13, Appl
4	55.5	24.4	553	2 US-08-288-065A-13	Sequence 13, Appl
5	55.5	24.4	553	2 US-08-362-240A-13	Sequence 13, Appl
6	55.5	24.4	553	3 US-08-477-459-14	Sequence 14, Appl
7	55.5	24.4	553	3 US-08-479-869-14	Sequence 14, Appl
8	55.5	24.4	553	4 US-08-486-414-14	Sequence 14, Appl
9	55.5	24.4	553	4 US-08-804-372A-11	Sequence 11, Appl
10	55.5	24.4	553	5 PCT-US94-01826A-14	Sequence 14, Appl
11	55.5	24.4	553	5 PCT-US94-02252A-14	Sequence 14, Appl
12	55.5	24.4	553	5 PCT-US95-10245-13	Sequence 13, Appl
13	55.5	24.4	553	6 5310678-1	Patent No. 5310678
14	55.5	24.4	837	4 US-09-564-805-228	Sequence 228, Appl
15	55.5	24.4	97	4 US-09-242-690A-6	Sequence 6, Appl
16	54	23.8	1912	1 US-08-409-995-4	Sequence 4, Appl
17	54	23.8	1912	1 US-08-685-467-4	Sequence 4, Appl
18	54	23.8	2353	4 US-09-377-155-33	Sequence 33, Appl
19	54	23.8	2353	4 US-08-913-942-4	Sequence 4, Appl
20	54	23.8	2353	4 US-09-669-974-33	Sequence 33, Appl
21	54	23.8	2354	4 US-09-268-347-47	Sequence 47, Appl
22	54	23.8	2411	4 US-09-268-347-36	Sequence 36, Appl
23	53	23.3	513	1 US-08-403-866-1	Sequence 1, Appl
24	51.5	22.7	244	4 US-08-919-573-2	Sequence 2, Appl
25	51.5	22.7	244	4 US-08-919-573-2	Sequence 2, Appl
26	49.5	21.8	436	3 US-08-486-099-94	Sequence 94, Appl
27	49.5	21.8	436	3 US-08-360-107A-104	Sequence 104, Appl

28	49.5	21.8	436	3 US-08-484-223B-94	Sequence 94, Appl
29	49.5	21.8	436	3 US-08-919-597-94	Sequence 94, Appl
30	49.5	21.8	436	3 US-08-475-668A-94	Sequence 94, Appl
31	49.5	21.8	436	3 US-08-485-551A-94	Sequence 94, Appl
32	49.5	21.8	436	3 US-08-472-913A-94	Sequence 94, Appl
33	49.5	21.8	436	4 US-08-485-264A-94	Sequence 94, Appl
34	49.5	21.8	436	4 US-08-474-349A-94	Sequence 94, Appl
35	48.5	21.4	202	4 US-08-858-207A-332	Sequence 48, Appl
36	48	21.1	720	3 US-09-257-799-48	Sequence 48, Appl
37	47.5	20.9	98	2 US-08-920-919A-48	Sequence 3, Appl
38	47.5	20.9	98	2 US-08-918-727-3	Sequence 3, Appl
39	47.5	20.9	98	3 US-09-205-680A-3	Sequence 3, Appl
40	47.5	20.9	98	3 US-08-948-889-11	Sequence 11, Appl
41	47.5	20.9	270	2 US-08-852-743-5	Sequence 5, Appl
42	47.5	20.9	270	3 US-09-185-370-5	Sequence 2, Appl
43	47.5	20.9	15281	2 US-08-471-119A-2	Sequence 53, Appl
44	47	20.7	92	3 US-08-813-884-53	Sequence 52, Appl
45	47	20.7	93	3 US-08-813-884-52	Sequence 52, Appl

ALIGNMENTS

RESULT 1
US-08-663-566A-13
Sequence 13, Application US/08663566A
Patent No. 5853733
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D
APPLICANT: Macdonald, Richard D
TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys
TITLE OF INVENTION: and Use Thereof
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,566A
FILING DATE: June 13, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400
TELEFAX: (212)391-0526
TELEX: 422523
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-663-566A-13

Query Match 24.4%; Score 55.5; DB 2; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

4 TSLKNDKFEKITDELEITGGGSLSTFRLNRSFTQALGK 46
DB 433 TRLSGEDVTVYOKNISIDSOVITITGMDISTELGNVNSISNALNK 480

RESULT 5
US-08-362-240A-13
Sequence 13, Application US/08362240A
Patent No. 5965138
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D
APPLICANT: Junker, David A
APPLICANT: Wild, Martha A
TITLE OF INVENTION: Recombinant Herpesvirus and Uses Thereof
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,240A
FILING DATE: Dec-22-94
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400
TELEFAX: (212)391-0526
TELEX: 422523
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-240A-13

Query Match 24.4%; Score 55.5; DB 2; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4; Mismatches 7; Indels 20; Gaps 5; Cuts 1;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

OY 4 TSLKND-----KEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TRLSGFDVTVOKNISIDSOVIITGNLDISTELGNVNSISNALNK 480

RESULT 6
US-08-477-459-14
Sequence 14, Application US/08477459
Patent No. 6001369
GENERAL INFORMATION:
APPLICANT: Mark D. Cochran
TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,459
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-459-14

Query Match 24.4%; Score 55.5; DB 3; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4; Mismatches 7; Indels 20; Gaps 5; Cuts 1;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

OY 4 TSLKND-----KEIKTDELEIIIGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TRLSGFDVTVOKNISIDSOVIITGNLDISTELGNVNSISNALNK 480

RESULT 7
US-08-479-869-14
Sequence 14, Application US/08479869
Patent No. 6123949
GENERAL INFORMATION:
APPLICANT: Cochran Ph.D. Mark D
TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,869
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/024,156
FILING DATE: 26-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-479-869-14

Query Match 24.4%; Score 55.5; DB 3; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4; Mismatches 7; Indels 20; Gaps 5; Cuts 1;

STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/02252A
FILING DATE: 28-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq, John P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)977-9550
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-02252A-14

Query Match 24.4%; Score 55.5; DB 5; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TLKNDP-----KEIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TLRISGEFDVYQKNISIDSOVYITGNLDISTELGNVNSISNALNK 480

RESULT 12
PCT-US95-10245-13

Sequence 13, Application PC/TUS9510245
GENERAL INFORMATION:
APPLICANT: SYNPRO CORPORATION

TITLE OF INVENTION: Recombinant Herpesvirus of Turkeys And Uses Thereof
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White

STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10245
FILING DATE: 09-AUG-1995
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: White, John P
REGISTRATION NUMBER: 28,678

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)278-0400
TELEFAX: (212)391-0526
TELEX: 422523

INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-10245-13

Query Match 24.4%; Score 55.5; DB 5; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TLKNDP-----KEIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TLRISGEFDVYQKNISIDSOVYITGNLDISTELGNVNSISNALNK 480

RESULT 13
5310678-1

Patent No. 5310678
APPLICANT: Bingham, Richard W.; Chambers, Phillip; Emerson, Peter
T. Miller, Neil S.

TITLE OF INVENTION: NEWCASTLE DISEASE VIRUS GENE CLONES
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/438,945

FILING DATE: 17-NOV-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 885,765

FILING DATE: 15-JUL-1986
SEQ ID NO: 1:
LENGTH: 553

5310678-1

Query Match 24.4%; Score 55.5; DB 6; Length 553;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TLKNDP-----KEIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
DB 433 TLRISGEFDVYQKNISIDSOVYITGNLDISTELGNVNSISNALNK 480

RESULT 14
US-09-564-805-228

Sequence 228, Application US/09564805
Patent No. 6333403
GENERAL INFORMATION:

APPLICANT: Tavligian, Sean V.
APPLICANT: Teng, David H.F.
APPLICANT: Simard, Jacques

APPLICANT: Rommens, Johanna M.
APPLICANT: Myriad Genetics, Inc.

TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
FILE REFERENCE: 2318-258
CURRENT APPLICATION NUMBER: US/09/564,805

CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/107,468
PRIOR FILING DATE: 1998-11-06

PRIOR APPLICATION NUMBER: 09/434,382
PRIOR FILING DATE: 1999-11-05
NUMBER OF SEQ ID NOS: 240

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 228
LENGTH: 837

TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-564-805-228

Query Match 24.4%; Score 55.5; DB 4; Length 837;
Best Local Similarity 34.1%; Pred. No. 5.8;
Matches 14; Conservative 9; Mismatches 9; Indels 9; Gaps 1;

QY 7 LKNDKEIKTDELEIIGSGS-----LSTFFRLFNK 38
DB 473 LPNCEKIRDMDEIVILIGSSGSPSKYRNVAIFIDLSR 513

RESULT 15
US-09-242-690A-6
; Sequence 6, Application US/09242690A
; Patent No. 6284534
; GENERAL INFORMATION:
; APPLICANT: KONDO, KEIJI
; APPLICANT: MIURA, YUTAKA
; TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
; TITLE OF INVENTION: SAME
; FILE REFERENCE: 049441/0118
; CURRENT APPLICATION NUMBER: US/09/242,690A
; CURRENT FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/JP97/02924
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: JP 8/241062
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Candida utilis
US-09-242-690A-6

Query Match 23.8%; Score 54; DB 4; Length 97;
Best Local Similarity 41.48; Pred. No. 0.59;
Matches 12; Conservative 6; Mismatches 11; Indels 0; Gaps 0;
Qy 1 MKKTLKNDKFKIKTDELEIIIGSGSL 29
|||: ||: ||| | : : |
Db 43 MKKTIENGFREIKGYEQLVVYASDKL 71

Search completed: July 30, 2002, 15:14:41
Job time: 517 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:08:49 ; Search time 29.21 Seconds
(without alignments)
151.322 Million cell updates/sec

Title: US-09-833-017-2

Perfect score: 227
Sequence: 1 MKKTLKNDPKEIKTDELE.....GSLSTPFRLFNRSFTQALGK 46

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	30.4	850	2	P95260 ABC transporter, p
2	69	30.4	850	2	A98126 conserved hypochet
3	60.5	26.7	475	2	S52893 hypothetical prote
4	58.5	25.8	41	2	H98126 competence stimula
5	58	25.6	1009	2	E99792 hypothetical prote
6	57.5	23.3	516	2	T32611 glucose-6-phosphat
7	56.5	24.9	398	2	AC3578 glycine betaine/L-
8	56.5	24.9	408	2	D86806 betaine ABC transp
9	56.5	24.9	593	2	C97848 ABC transporter AT
10	56.5	24.9	893	2	S64926 probable membrane
11	56	24.7	719	2	T12258 pyroline-5-carbox
12	56	24.7	1250	2	E81339 probable restricti
13	55.5	24.4	416	2	AC1143 probable NAD(P)-de
14	55.5	24.4	416	2	AC1501 probable NAD(P)-de
15	55.5	24.4	553	1	VGNZNV cell fusion glycop
16	55.5	24.4	553	1	VGNZNV cell fusion glycop
17	55.5	24.4	553	1	D46329 cell fusion glycop
18	55.5	24.4	553	1	E46329 cell fusion glycop
19	55.5	24.4	553	1	G46329 cell fusion glycop
20	55.5	24.4	553	1	P96561 unknown protein [l
21	55	24.2	278	2	AB3091 hypothetical prote
22	55	24.2	278	2	H98195 hypothetical prote
23	55	24.2	608	2	S51790 centromere protein
24	54.5	24.0	230	2	D86746 cell-division ATP-
25	54.5	24.0	361	2	D98243 opuna (AF234619) [
26	54.5	24.0	361	2	AB3042 histidinol-phospha
27	54.5	24.0	361	2	D64419 histidinol-phospha
28	54.5	24.0	463	2	D97814 chromosomal replic
29	54.5	24.0	720	2	S73267 phenylalanine--trn

30	54	23.8	226	1	C49769 gap junction prote
31	54	23.8	226	1	A33646 gap junction prote
32	54	23.8	243	2	C96946 glutamine ABC tra
33	54	23.8	708	2	T24727 hypothetical prote
34	54	23.8	1292	2	F64237 DNA-directed RNA p
35	53.5	23.6	165	2	S16722 coenzyme F420 hyd
36	53.5	23.6	241	2	AC1356 amino acid ABC tra
37	53.5	23.6	241	2	AD1726 amino acid ABC tra
38	53.5	23.6	242	2	AH1180 amino acid ABC tra
39	53.5	23.6	242	2	A11537 amino acid ABC tra
40	53.5	23.6	243	2	D97008 ABC-type polar ami
41	53.5	23.6	246	2	G81409 probable phosphate
42	53.5	23.6	331	2	H81300 probable anion-upt
43	53.5	23.6	553	1	VGNZ01 cell fusion glycop
44	53.5	23.6	553	1	A46329 cell fusion glycop
45	53.5	23.6	553	1	B46329 cell fusion glycop

ALIGNMENTS

```
RESULT 1
F95260
ABC transporter, permease protein; probable SP2231 [Imported] - Streptococcus pneumon
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: F95260
R:Retelien, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-850 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76279.1; PID:gl4973742; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP2231

Query Match 30.4%; Score 69; DB 2; Length 850;
Best Local Similarity 33.3%; Pred. No. 1.1;
Matches 19; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

QY 2 KKTLSLKNDP-----KEIKTDELEITIGSGSLG-----TFRRLNRSFTQALGK 46
DB 692 KKTFTDNVFSFENLGYTKERKKNINVSFGNSQVSPESPTFYRLDTKTFTEAIQK 748

RESULT 2
A98126
Conserved hypothetical protein spr2036 [Imported] - Streptococcus pneumoniae (strain
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: A98126
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain K6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: A98126
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-850 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAU00838.1; PID:gl5459743; GSPDB:GN00174
C:Genetics:
A:Gene: spr2036
```

```

Query Match      30.4%  Score 69;  DB 2;  Length 850;
Best Local Similarity 33.3%  Pred. No. 1.1;
Matches 19;  Conservative 10;  Mismatches 16;  Indels 12;  Gaps 2;

QY      2  KRTSLKNDP-----KEIKTDELEIIIGSGSL-----TFRLRNRSTQALGK 46
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      692  KRTFTDWNVFFSNLGYKEKKTFNNINVSFPGNSQVSFSEPTFYRLDKTFTTEAIQK 748

RESULT      3
S52893
hypothetical protein YMR044w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YM9352.09
C:Species: Saccharomyces cerevisiae
C:Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 05-Nov-1999
C:Accession: S52893
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52885
A:Accession: S52893
A:Molecule type: DNA
A:Crosses: 1-475 <ODE>
A:Cross-references: EMBL:Z48502; NID:g695715; PIDN:CAA88410.1; PID:g695724; GSPDB:GNO0001
C:Genetics:
A:Gene: MIPS:YMR044w
A:Map position: 138k

```

Query Match	26.7%	Score 60.5	DB 2	length 475
Best Local Similarity	38.9%	Pred. No. 6.9		
Matches 14	Conservative 9	Mismatches 12	Indels 1	Gaps 1

QY	1	MKKTUSLAKNDEKREKTDELEIIIGCGSLSTFFRLE	36
:	:	: : : : : :	:
DB	361	INQAVSIRKDF--ELIIDEQLIALDTRGSRNEETITTF	395

RESULT 4
 H98126
 competence stimulating peptide precursor (CSP) [imported] - Streptococcus pneumoniae (strain H98126)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C:Accession: H98126
 R:Host(s): J.A. Alborn Jr., W. Arnold, J. Blaszczak, L. Burgett, S. Dehoff, B.S. Dyer, R. Leblanc, D.J. Lee, L.N. Letkowitz, E.J. Lu, J. Matsushima, P. McAhren, S. Meyer, P. Sun, P.M. Winkler, M.E.
 Y. R. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
 A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: H98126
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 141 <KIR>
 A:Cross-references: GB:AE007317; PIDN:AAI00845.1; PID:g15459751; GSPDB:GN00174
 C:Genetics:
 A:Gene: comC

		Query Match	25.8%;	Score 58.5;	DB 2;	length 41;
		Best Local Similarity	44.4%;	Pred. No. 0.83;		
		Matches	16;	Conservative	5;	Mismatches 14; Indels 1; Gaps 1.
OY	1	MKKTILSKDNDFKEIKDELEIIIGSSSLSTFRRLF	36			
		: : - : : :				
Db	1	MKNVTVKLE-QFVALKEQDLQRIKGGEEMRLSIFRNF	35			

RESULT	5
E89792	
hypothetical protein SA0272 [imported] - Staphylococcus aureus (strain N315)	

C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: E89792
R:Kunoda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratazu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: E89792
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1009 <KUR>
A:Cross-References: GB:BA000018; PTD:913700198; PIDN:BA841496.1; GSPDB:GN00149
A:Experimental source: strain N15
C:Genetics:
A:Gene: SA0272

Query Match	25.6%	Score 58	DB 2	Length 1009
Best Local Similarity	34.3%	Pred. NO. 34		
Matches	12	Conservative	8	Mismatches 15; Indels 0; Gaps 0;
QY	9	NDFKEKTDLELEITIGSGSLSTFFRLERSTQA	43	
DB	202	NDFPELEFDTLVNSISANKDITKMFOTYKSLISA	236	

RESULT 6
T52611
glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) [imported] - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52611
R:Went, U.K.; Haenschild, R.; Lange, C.; Pietersema, M.; Wenderoth, I.; von Schaewen,
Plant Mol. Biol. 40, 487-494, 1999
A:Title: Evidence for functional convergence of redox regulation in G6PDH isoforms of
A:Reference number: Z25284
A:Accession: T52611
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-516 <MEN>
A:Cross-references: EMBL:AJ010970; PIDN:CAB52674.1
C:Genetics:
A:Gene: acg9
C:Superfamily: glucose-6-phosphate dehydrogenase
C:Keywords: oxidoreductase

	Query Match	25.3%	Score 57.5;	DB 2;	Length 516;
	Best Local Similarity	41.5%;	Pred. No. 19;		
	Matches	22;	Conservative	7;	Mismatches 11; Indels 13; Gaps 5;
OY	1 MKTLLSLKND-EKE-----IKTDELEII-GGSGSIS---TF---FRINRSP	40			
	: : :				
Dd	8 MEKSTLTKNDSFVKEYNPIETGTSLSTLVAGASDGLAKKTFFALTNHRCF	60			

RESULT 7
AC3578
glycine betaine/1-proline transport ATP-binding protein *prov* BMEII0548 [imported] - B
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence-revision 01-Feb-2002 #text-change 01-Feb-2002
C:Accession: AC3578
R:DelVecchio, V.C.; Kapatral, V.; Kapatral, R.J.; Patra, G.; Muter, C.; Los, T.; Ivanov
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A:Reference number: AD3252; PMID:1175668
A:Accession: AC3578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <KUR>

A:Cross-references: GB:AE008918; PIDN:AAL53790.1; PID:g17984720; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0548
A:Map position: II

Query Match 24.9%; Score 56.5; DB 2; Length 398;
Best Local Similarity 37.2%; Pred. No. 19;
Matches 16; Conservative 7; Mismatches 19; Indels 1; Gaps 1;

OY 4 TLSLKNDKEIKTDELEIIIGSGS-LSTFFRLFNRSFTQALG 45
DB 55 TIGVDNATFDRREGFVIMGLSSGSKSTLRLLNRLIEPTAG 97

RESULT 8

D86806
betaine ABC transporter ATP binding protein busA [imported] - Lactococcus lactis subsp.
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: D86806
R:Botolin, A.; Whicker, P.; Mauge, S.; Jaillon, O.; Malarme, K.; Weissbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86825; MUID:11235186; PMID:11337471
A:Accession: D86806
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-408 <STO>
A:Cross-references: GB:AE005176; PID:g12724443; PIDN:AAK05550.1; GSPDB:GN00146
A:Experimental source: strain IL403
C:Genetics:
A:Gene: busA

Query Match 24.9%; Score 56.5; DB 2; Length 408;
Best Local Similarity 45.7%; Pred. No. 19;
Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 13 EIKTDELEIIIGSGS-LSTFFRLFNRSFTQALG 46
DB 50 EINEGEIVIMGLSSGSKSTLRLLNRLIEPTSGK 84

RESULT 9

C97848
ABC transporter ATP-binding protein uup [imported] - Rickettsia conorii (strain Malish 7
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: C97848
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: C97848
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-593 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03725.1; PID:g15620316; GSPDB:GN00173
C:Genetics:
A:Gene: uup
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 24.9%; Score 56.5; DB 2; Length 593;
Best Local Similarity 37.0%; Pred. No. 29;
Matches 17; Conservative 5; Mismatches 23; Indels 1; Gaps 1;

OY 2 KKTSLKNDKEIKTDELEIIIGSGS-LSTFFRLFNRSFTQALG 46
DB 298 KMTKINNFRVKNKGKIGIANGSGSKSTFIRLLTKQLTPESGK 343

RESULT 10

S64926
probable membrane protein YLR092W - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein L2528
C:Species: Saccharomyces cerevisiae
C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 06-Feb-1998
C:Accession: S64926
R:Benes, V.; Reichtman, S.; Nentwich, U.; Schwager, C.; Ansoorge, W.; Voss, H.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64920
A:Accession: S64926
A:Molecule type: DNA
A:Residues: 1-893 <BEN>
A:Cross-references: EMBL:Z73264; NID:g1360466; PID:e245799; PID:g1360467; MIPS:YLR092
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SCD:YLR092W
A:Cross-references: SCD:S0004082; MIPS:YLR092W
A:Map position: 12R

C:Keywords: transmembrane protein
F:136-152/Domain: transmembrane #status predicted <TM1>
F:223-239/Domain: transmembrane #status predicted <TM2>
F:244-260/Domain: transmembrane #status predicted <TM3>
F:302-318/Domain: transmembrane #status predicted <TM4>
F:358-374/Domain: transmembrane #status predicted <TM5>
F:411-427/Domain: transmembrane #status predicted <TM6>
F:488-504/Domain: transmembrane #status predicted <TM7>
F:538-554/Domain: transmembrane #status predicted <TM8>
F:557-573/Domain: transmembrane #status predicted <TM9>

Query Match 24.9%; Score 56.5; DB 2; Length 893;
Best Local Similarity 36.7%; Pred. No. 46;
Matches 18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;

OY 1 MKKTSLKNDKEIKTDELEIIIGSGSLSTFFRL-----FKRSTQA 43
DB 426 ISKSPGRINDYKVPDDEL-IAIGVSNLGTFFNAYPATGSPRSALKA 473

RESULT 11

T12258
pyrroline-5-carboxylate synthetase (PC 1.5.1.-) - common ice plant
C:Species: Mesembryanthemum crystallinum (common ice plant)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 17-Nov-2000
C:Accession: T12258
R:Michalowski, C.B.; Quigley-Landreau, F.; Bohner, H.J.
submitted to the EMBL data library, May 1998
A:Description: Mesembryanthemum crystallinum pyrroline-5-carboxylate synthetase mRNA.
A:Reference number: Z17475
A:Accession: Y12258
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-719 <MTC>
A:Cross-references: EMBL:AF067967; NID:g3176964; PID:g3176965
C:Genetics:
A:Gene: P5CS
C:Superfamily: glutamate-5-semialdehyde dehydrogenase
C:Keywords: oxidoreductase

Query Match 24.7%; Score 56; DB 2; Length 719;
Best Local Similarity 36.6%; Pred. No. 42;
Matches 15; Conservative 7; Mismatches 13; Indels 6; Gaps 1;

OY 6 SLKNDKEIKTDELEIIIGSGSLST-----PFRFNRF 40
DB 37 SLCEOLKELNSDGYEVILVTSAGVAGRORLRFKRLVNSSF 77

RESULT 12

B81339

Wed Jul 31 08:30:53 2002

us-09-833-017-2.rpr

Page 5

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:14:49 ; Search time 15.86 Seconds

(Without alignments)
112.301 Million cell updates/sec

Title: US-09-833-017-2

Sequence: 1 MKRTSLKNDKFEIKTDELE.....GSLSTFFRLFNRSFTQALGK 46

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	60.5	26.7	475 1 YMTA_YEAST	004213 saccharomyc
2	58.5	25.8	41 1 CSPL_STRPN	054712 streptococ
3	56.5	24.9	893 1 SOL2_YEAST	012325 saccharomyc
4	56	24.7	719 1 PSCS_MESCR	065361 m delta 1-p
5	55.5	24.4	41 1 CSP_STROR	033689 streptococ
6	55.5	24.4	553 1 VGLF_NDV4	P06196 newcastle d
7	55.5	24.4	553 1 VGLF_NDV4	P33613 newcastle d
8	55.5	24.4	553 1 VGLF_NDV4	P33614 newcastle d
9	55.5	24.4	553 1 VGLF_NDV4	P12571 newcastle d
10	55	24.2	608 1 CB32_YEAST	P40969 saccharomyc
11	54.5	24.0	373 1 HIS8_METUA	Q58365 methanococ
12	54.5	24.0	463 1 DNMA_RICCN	Q92856 porphyra pu
13	54.5	24.0	720 1 SYFB_PORPU	P51346 porphyra pu
14	54	23.8	226 1 CXB2_MOUSE	000977 mus musculu
15	54	23.8	226 1 CXB2_MOUSE	P21994 rattus norv
16	54	23.8	708 1 YNZB_CARFL	P45972 caenorhabdi
17	54	23.8	1292 1 RPOC_MYCGE	P47582 mycoplasma
18	53.5	23.6	259 1 PSTB_EDMTA	Q9am14 Edwardsiell
19	53.5	23.6	553 1 VGLF_NDV4	P33936 newcastle d
20	53.5	23.6	553 1 VGLF_NDV4	P33615 newcastle d
21	53.5	23.6	553 1 VGLF_NDV4	P12570 newcastle d
22	53.5	23.6	553 1 VGLF_NDV4	P12570 newcastle d
23	53	23.3	274 1 PPNK_AQUAE	067055 aquifex aeo
24	53	23.3	400 1 PROV_ECOLI	P14175 escherichia
25	53	23.3	400 1 PROV_ECOLI	P17328 salmonella
26	53	23.3	513 1 LEU1_YEAST	002141 lactococcus
27	53	23.3	513 1 LEU1_YEAST	002141 lactococcus
28	53	23.3	1132 1 DNBI_HSV6U	P52538 human herpe
29	53	23.3	1132 1 DNBI_HSV6U	P52538 human herpe
30	52.5	23.1	310 1 GPRB_BP22	P57022 bacillopho
31	52.5	23.1	564 1 V61A_MYCPN	P75041 mycoplasma
32	51.5	22.7	41 1 CSP2_STROR	033689 streptococ
33	51.5	22.7	240 1 PPRH_METUA	058656 methanococ

34	51.5	22.7	252 1 Y830_METUA	Q58240 methanococ
35	51.5	22.7	257 1 PSTB_ECOLI	P07655 escherichia
36	51.5	22.7	259 1 DNMA_MYCBO	P49991 mycobacteri
37	51.5	22.7	306 1 GTRB_ECOLI	P77293 escherichia
38	51	22.5	306 1 GTRB_ECOLI	P03866 staphylococ
39	51	22.5	295 1 UNG_MOUSE	P97931 mus musculu
40	51	22.5	304 1 UNG_MOUSE	P13051 homo sapien
41	51	22.5	360 1 DFERA_DIAA	P51104 dianthus ca
42	51	22.5	397 1 Y636_METUA	Q58053 methanococ
43	51	22.5	461 1 Y061_MYCGE	P47307 mycoplasma
44	51	22.5	465 1 CG12_CANAL	P43062 candida alb
45	51	22.5	511 1 G6PD_EMBNT	P41764 emeticella

ALIGNMENTS

RESULT 1
YMTA_YEAST STANDARD; PRT; 475 AA.

AC Q04213;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 55.4 kDa protein in MCM1-NUP116 intergenic region.
GN YMR044W OR YMR532.09.
OS Saccharomycetes cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL; Z48502; CAA8410.1;
DR SGD; S0004647; YMR044W.
DR InterPro: IPR000313; PMWP.
DR Pfam: PF00855; PMWP; 1.
DR SMART; SM00293; PMWP; 1.
KW Hypothetical protein.
FT DOMAIN 74 77 POLY-GLU.
FT DOMAIN 186 195 POLY-GLU.
FT DOMAIN 243 251 POLY-GLU.
SQ SEQUENCE 475 AA; 55426 MW; DFD618981178060 CRC64;

Query Match 26.7%; Score 60.5; DB 1; Length 475;
Best Local Similarity 38.9%; Pred. No. 2.8;
Matches 14; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

OY 1 MKRTSLKNDKFEIKTDELEITIGSGSLSTFFRLF 36
DB 361 INQAVSINKDF-ETILDELQIALDTRGSRNFEITIF 395

RESULT 2
CSPL_STRPN STANDARD; PRT; 41 AA.

AC Q54712;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Competence stimulating peptide type I precursor (CSP-1).

CC COMC1 OR COMC.
 OS Streptococcus pneumoniae.
 CC Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 CC NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 25-41.
 RC STRAIN=RX / Cpl200;
 RX MEDLINE=96074663; PubMed=7479953;
 RA Haavarsen L.S., Coomaraswamy G., Morrison D.A.;
 RT "An unmodified heptadecapeptide pheromone induces competence for
 genetic transformation in Streptococcus pneumoniae.";
 RT Proc. Natl. Acad. Sci. U.S.A. 92:11140-11144(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RX;
 RX MEDLINE=97206147; PubMed=9157240;
 RA Cheng O., Campbell E.A., Naughton A.M., Johnson S., Mezure H.R.;
 RT "The *com* locus controls genetic transformation in Streptococcus
 pneumoniae.";
 RT Mol. Microbiol. 23:683-692(1997).
 CC -1- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
 FOR GENETIC TRANSFORMATION.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE COMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U33315; AAC44440.1; -;
 DR EMBL: U76218; AAC44895.1; -;
 DR InterPro: IPR004286; Comc.
 DR Pfam: PF03047; Comc; 1.
 DR Phospho: Competence.
 FT PROPEP 1 24
 FT CHAIN 25 41
 FT SEQUENCE 41 AA; 4971 MW; 1448B414E980E86A CRC64;
 SQ
 Query Match 25.8%; Score 58.5; DB 1; Length 41;
 Best Local Similarity 44.4%; Pred. No. 0.34;
 Matches 16; Conservative 5; Mismatches 14; Indels 1; Gaps 1;
 OY 1 MKKTLISKNDPKREIKDELEITIGSGSLSTPRL 36
 DB 1 MKNTVKLE-QFVALKEKDLOKIKGEMRLSKFTRDF 35
 RESULT 3
 SUL2_YEAST
 ID SUL2_YEAST STANDARD; PRT; 893 AA.
 AC Q12325;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Sulfate permease 2 (High-affinity sulfate transporter 2).
 GN SUL2 OR SEL2 OR YLR092W OR L9449.1.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gatung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnson L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,

RA Miller N., Nhan M., Pauley A., Peluso D., Rifkin L., Riles L.,
 RA Tach A., Trevasaki E., Vignati D., Wilcox L., Woldman P., Vandin M.,
 RA Wilson R., Waterston R.;
 RN Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Wentlich U., Schwager C., Ansoerge W., Voss H.;
 RN Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=97207837; PubMed=9055073;
 RA Cherest H., Davidian J.C., Thomas D., Benes V., Ansoerge W.,
 RA Surdita-Kerjan Y.;
 RT "Molecular characterization of two high affinity sulfate transporters
 in Saccharomyces cerevisiae.";
 RT Genetics 145:627-635(1997).
 CC -1- FUNCTION: HIGH AFFINITY UPTAKE OF SULFATE INTO THE CELL.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U53880; AAB67596.1; -;
 DR EMBL: Z73264; CAA97653.1; -;
 DR EMBL: Z73265; CAA97655.1; -;
 DR EMBL: U53876; AAB67550.1; -;
 DR SGD: S0004082; SUL2.
 DR InterPro: IPR002645; STAS.
 DR InterPro: IPR001902; Sulfate_transp.
 DR Pfam: PF01740; STAS; 1.
 DR Pfam: PF00916; Sulfate_transp; 1.
 DR PROSITE: PS01130; SULFATE_TRANS; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT TRANSMEM 132 152
 FT TRANSMEM 164 184
 FT TRANSMEM 189 209
 FT TRANSMEM 222 242
 FT TRANSMEM 245 265
 FT TRANSMEM 306 326
 FT TRANSMEM 351 371
 FT TRANSMEM 400 420
 FT TRANSMEM 444 464
 FT TRANSMEM 484 504
 FT TRANSMEM 539 559
 FT TRANSMEM 560 580
 FT SEQUENCE 893 AA; 99650 MW; 67826955AC7C0BF5 CRC64;
 SQ
 Query Match 24.9%; Score 56.5; DB 1; Length 893;
 Best Local Similarity 36.7%; Pred. No. 19;
 Matches 18; Conservative 6; Mismatches 18; Indels 7; Gaps 2;
 OY 1 MKKTLISKNDPKREIKDELEITIGSGSLSTPRL-----FNRSFTQA 43
 DB 426 ISMSFGRINDYKVPDDEL-IALGVNLTGFNAYPATGSPRSALKA 473
 RESULT 4
 P5CS_MESCR
 ID P5CS_MESCR STANDARD; PRT; 719 AA.
 AC O65361;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Delta 1-pyrroline-5-carboxylate synthetase (P5CS) [includes: Glutamate
 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK); Gamma-glutamyl
 DE phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde

DE dehydrogenase) (glutamyl-gamma-semialdehyde dehydrogenase)].
 GN P5CS
 OS Mesembryanthemum crystallinum (Common ice plant).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Caryophyllales; Alzaceae; Mesembryanthemum.
 NCBI_TaxID=3544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Michalowski C.B., Ouygley-Landreau F., Bohnert H.J.;
 RT Mesembryanthemum crystallinum pyruvate-5-carboxylate synthetase
 RT mRNA.
 RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: P5CS PLAYS A KEY ROLE IN PROLINE BIOSYNTHESIS, LEADING
 CC TO OSMOREGULATION IN PLANTS.
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
 CC phosphate.
 CC -1- CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate +
 CC NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.
 CC -1- ENZYME REGULATION: FEEDBACK REGULATED BY PROLINE.
 CC -1- PATHWAY: FIRST AND SECOND STEPS IN PROLINE BIOSYNTHESIS PATHWAY.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN LEAVES AND IS
 CC INDUCIBLE IN ROOTS SUBJECTED TO SALT STRESS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GLUTAMATE 5-
 CC KINASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GAMMA-
 CC GLUTAMYL PHOSPHATE REDUCTASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF067967; AAC18862.1;
 DR Mendel: 29718; Meser:1229;29718.
 DR InterPro: IPR001048; AakInase.
 DR InterPro: IPR002086; Aldehyde_dehydr.
 DR InterPro: IPR000965; GPR.
 DR InterPro: IPR001057; Glut_5_kinase.
 DR Pfam: PF00696; aakInase; 1.
 DR Pfam: PF00171; aldehyd; 1.
 DR PRINTS: PR00474; GLUTAMATE_5_KINASE; 1.
 DR PROSITE: PS00902; GLUTAMATE_5_KINASE; 1.
 DR PROSITE: PS01223; PROA; 1.
 KM Proline biosynthesis; Multifunctional enzyme; Oxidoreductase; NADP;
 KW transferase; kinase.
 FT DOMAIN 1 293 GLUTAMATE 5-KINASE.
 FT DOMAIN 294 719 GAMMA-GLUTAMYL PHOSPHATE REDUCTASE.
 FT SEQUENCE 719 AA: 77855 MW: 1510D1AF5559961 CRC64;
 SQ
 Query Match 24.7%; Score 56; DB 1; Length 719;
 Best Local Similarity 36.6%; Pred. No. 17;
 Matches 15; Conservative 7; Mismatches 13; Indels 6; Gaps 1;
 QY 6 SLKNDPEIKTDELEIIIGSGSLST-----FFRLKRSF 40
 DB 37 SLCEQLKEINSDGVEYIVLTGSAVSAGRQRLFRKLVNSSF 77
 RESULT 5
 CSP_STROR STANDARD: PRT; 41 AA.
 AC 033689;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Competence stimulating peptide precursor (CSP).
 GN COMC.
 OS Streptococcus oralis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OC NCBI_TaxID=1303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIC 11427;
 RX MEDLINE=98012953; PubMed=9352904;
 RA Haeverstein L.S., Hakenbeck R., Gaustad P.;
 RT "Natural competence in the genus Streptococcus: evidence that
 RT streptococci can change phenotype by interspecies recombinational
 RT exchanges".
 RL J. Bacteriol. 179:6589-6594(1997).
 CC -1- FUNCTION: ACTS AS A PHEROMONE, INDUCES CELLS TO DEVELOP COMPETENCE
 CC FOR GENETIC TRANSFORMATION.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE COMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ000873; CA04363.1; -
 DR InterPro: IPR004288; COMC.
 DR Pfam: PF03047; COMC; 1.
 KW Pheromone; Competence.
 FT PROPEP 1 24 POTENTIAL.
 FT CHAIN 25 41 COMPETENCE STIMULATING PEPTIDE.
 SQ SEQUENCE 41 AA: 4988 MW: 127E5B6F72A1463 CRC64;
 QY 1 MKKTSLKNDPEIKTDELEIIIGSGSLSTFFR-LF-NRS 39
 DB 1 MKNTKLE-QRFKEVTEALEIGRSGDKRLPYFFKILPSNRT 40
 RESULT 6
 ID VGLF_NDVV STANDARD: PRT; 553 AA.
 AC P06156;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2;
 DE Fusion glycoprotein F1].
 GN F.
 OS Newcastle disease virus (strain Beaudette C/45) (NDV).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OC NCBI_TaxID=11178;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87085486; PubMed=3025345;
 RA Chambers P., Millar N.S., Emmerson P.T.;
 RT "Nucleotide sequence of the gene encoding the fusion glycoprotein of
 RT Newcastle disease virus".
 RL J. Gen. Virol. 67:2685-2694(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89204698; PubMed=2705298;
 RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
 RA Nagai Y.;
 RT "Newcastle disease virus evolution. II. Lack of gene recombination in
 RT generating virulent and avirulent strains".
 RL Virology 169:273-282(1989).
 CC -1- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR

```
CC MEMBRANES.
CC -1- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: X04719; CAA28426.1; -
DR EMBL: M24697; AAA46648.1; -
DR EMBL: A03663; CAA00288.1; -
DR PIR: A27008; VGNZNY.
DR PIR: F46329; F46329.
DR HSSP: P04849; 1SVF.
DR InterPro: IPR00776; Fusion_gly.
DR Pfam: PF00523; Fusion_gly; 1.
KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 553 FUSION GLYCOPROTEIN F0.
FT CHAIN 26 116 F2 PROTEIN.
FT CHAIN 117 553 F1 PROTEIN.
FT TRANSMEM 117 136 POTENTIAL.
FT DOMAIN 137 500 EXTRACELLULAR.
FT TRANSMEM 501 527 POTENTIAL.
FT DOMAIN 527 553 CYTOPLASMIC.
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 523 523 PALMITATE (POTENTIAL).
SQ SEQUENCE 553 AA; 59041 MW; 63BFD1692AFDE191 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;
Best Local Similarity 33.3%; Pred. No. 15;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TSLKNDP-----KEIKTDELEIIIGSGSLSTFFRLNRSFTQALGK 46
DB 433 TRLSGEPDVTYQKNISIDQSVIITGNLDISTELGNVNSISNALNK 480

RESULT 7
VGLF_NDVH4 STANDARD; PRT; 553 AA.
ID VGLF_NDVH4
AC P33613;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fusion glycoprotein precursor [contains: Fusion glycoprotein F2;
DE Fusion glycoprotein F1].
GN F.
OS Newcastle disease virus (strain BI-Hitchner/47) (NDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirinae.
OX NCBI_TaxID=11181;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204898; PubMed=2705298;
RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA Nagai Y.;
RT "Newcastle disease virus evolution. II. Lack of gene recombination in
RT generating virulent and avirulent strains.";
RL Virology 169:273-282(1989).
CC -1- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
```

```
CC MEMBRANES.
CC -1- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES FUSION GLYCOPROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL: M24695; AAA46646.1; -
DR PIR: D46329; D46329.
DR HSSP: P04849; 1SVF.
DR InterPro: IPR00776; Fusion_gly.
DR Pfam: PF00523; Fusion_gly; 1.
KW Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal;
KW Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 553 FUSION GLYCOPROTEIN F0.
FT CHAIN 26 116 F2 PROTEIN.
FT CHAIN 117 553 F1 PROTEIN.
FT TRANSMEM 117 136 POTENTIAL.
FT DOMAIN 137 500 EXTRACELLULAR.
FT TRANSMEM 501 527 POTENTIAL.
FT DOMAIN 527 553 CYTOPLASMIC.
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 447 447 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT LIPID 523 523 PALMITATE (POTENTIAL).
SQ SEQUENCE 553 AA; 59909 MW; F163B9B/CPEBA493 CRC64;

Query Match 24.4%; Score 55.5; DB 1; Length 553;
Best Local Similarity 33.3%; Pred. No. 15;
Matches 16; Conservative 7; Mismatches 20; Indels 5; Gaps 1;

QY 4 TSLKNDP-----KEIKTDELEIIIGSGSLSTFFRLNRSFTQALGK 46
DB 433 TRLSGEPDVTYQKNISIDQSVIITGNLDISTELGNVNSISNALNK 480

RESULT 8
VGLF_NDVH4 STANDARD; PRT; 553 AA.
ID VGLF_NDVH4
AC P33614;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fusion glycoprotein precursor [contains: Fusion glycoprotein F2;
DE Fusion glycoprotein F1].
GN F.
OS Newcastle disease virus (strain Las/46) (NDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirinae.
OX NCBI_TaxID=11184;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89204898; PubMed=2705298;
RA Toyoda T., Sakaguchi T., Hirota H., Gotoh B., Kuma K., Miyata T.,
RA Nagai Y.;
RT "Newcastle disease virus evolution. II. Lack of gene recombination in
RT generating virulent and avirulent strains.";
RL Virology 169:273-282(1989).
CC -1- FUNCTION: THIS PROTEIN DIRECTS FUSION OF VIRAL AND CELLULAR
CC MEMBRANES.
CC -1- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2
CC LINKED BY A DISULFIDE BOND.
```


RL EMBL J. 13:5203-5211(1994).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95181532; PubMed-7876302;
 RA Strunskov A.V., Kingsbury J., Koshland D.;
 RT "CBF3 encodes a centromere protein of *Saccharomyces cerevisiae*";
 RL J. Cell Biol. 128:749-760(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RA Hunt S., Bowman S., Bartell B.G., Rejandream M.A.;
 RL Submitted (May-1995) to the EMBL/GenBank/DBD databases.
 CC -1- FUNCTION: ESSENTIAL FOR CHROMOSOME SEGREGATION AND MOVEMENT OF
 CC CENTROMERES ALONG MICROTUBULES. IT PLAYS A ROLE IN THE ATTACHMENT
 CC OF CHROMOSOMES TO THE SPINDLE. CBF3 BINDS SELECTIVELY TO A HIGHLY
 CC CONSERVED DNA SEQUENCE CALLED CDEIII, FOUND IN CENTROMERES AND IN
 CC SEVERAL PROMOTERS.
 CC -1- SUBUNIT: CBF3 IS FORMED OF FOUR SUBUNITS, CBF3A (CBF2), CBF3B
 CC (CBF3), CBF3C (CTF13) AND CBF3D.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINDCLEAR
 CC CLUSTER DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X81396; CAA57159.1; -
 DR EMBL: 012339; AAA57074.1; -
 DR EMBL: 249705; CAA89804.1; -
 DR SGD: S0004778; CBF3.
 DR InterPro: IPR001138; ZN2_Cy6_fungal.
 DR Pfam: PF00172; ZN2_Cy6_1.
 DR SMART: SM00066; GAL4; 1.
 DR PROSITE: PS00463; ZN2_Cy6_FUNGAL_1; 1.
 DR PROSITE: PS50048; ZN2_Cy6_FUNGAL_2; 1.
 KW Nuclear protein; Chromosomal protein; DNA-binding; Zinc;
 KW Metal-binding; Centromere.
 FT DNA BIND 14 42 ZN(2)-CYS(6); FUNGAL-TYPE.
 SQ SEQUENCE 608 AA; 71358 MW; 2E24A0508080A09B CRC64;

 Query Match 24.28; Score 55; DB 1; Length 608;
 Best Local Similarity 39.58; Pred. No. 19;
 Matches 15; Conservative 6; Mismatches 15; Indels 2; Gaps 1;

 QY 7 LKNDKFEIKTDELEIIGSGSLSTFFRLNRSFTQ 42
 DB 541 LQNDKRIEKLKNDKMFSLIKIGSLVPLNKRQESILE 578

 RESULT 11
 HIS8_METUA STANDARD; PRT; 373 AA.
 AC 058365;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable histidyl-lysine aminotransferase (EC 2.6.1.9) (Imidazole
 DE acetyl-phosphate transaminase).
 GN HIS8 OR M00935.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE-96337999; PubMed-8686087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, *Methanococcus*
 RT *jannaschii*";
 RL Science 273:1058-1073(1996).
 CC -1- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-
 CC (imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate (By
 CC SIMILARITY).
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 CC -1- PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.
 CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U67539; AB98960.1; -
 DR TIGR: M00935; -
 DR InterPro: IPR001511; Aminotran_1.
 DR InterPro: IPR001917; Aminotran_2.
 DR Pfam: PF00155; aminotran_1_2; 1.
 DR PROSITE: PS00599; AA_TRANSF_CLASS_2; FALSE_NEG.
 KW Histidine biosynthesis; Transferase; Aminotransferase;
 KW Pyridoxal phosphate; Complete proteome.
 FT BINDING 231 231 PYRIDOXAL PHOSPHATE (PROBABLE).
 SQ SEQUENCE 373 AA; 42961 MW; 369DB5EB4C671218 CRC64;

 Query Match 24.08; Score 54.5; DB 1; Length 373;
 Best Local Similarity 39.48; Pred. No. 13;
 Matches 13; Conservative 5; Mismatches 12; Indels 3; Gaps 1;

 QY 7 LKNDKFEIKTDELEIIGSGS---LSTFRLRF 36
 DB 79 MKELSKFLVDEENILVGGGDADELIDIRRTF 111

 RESULT 12
 DNAA_RICCN STANDARD; PRT; 463 AA.
 AC 092856;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Chromosomal replication initiator protein dnna.
 GN DNAA OR RC00916.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Malish 7;
 RX MEDLINE-21442074; PubMed-11557893;
 RA Ogata H., Audic S., Renseto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*";
 RL Science 293:2093-2098(2001).
 CC -1- FUNCTION: Plays an important role in the initiation and regulation
 CC of chromosomal replication. Binds to the origin of replication. It
 CC binds specifically double-stranded DNA at a 9 bp consensus (dnna)

CC box): 5'-TATAC(C/A)(C/A)-3'. Dnaa binds to ATP and to acidic
 CC phospholipids (BY similarity).
 CC -1- SIMILARITY: BELONGS TO THE DNAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE008645; AAL03454.1;
 CC InterPro: IPR001957; Bac_Dnaa.
 CC Pfam: PF00308; bac_dnaa; 1.
 CC PROSITE: PS01008; DNAa; 1.
 CC DNA replication; DNA-binding; ATP-binding; Complete proteome.
 CC NP_BIND 168 175 ATP (POTENTIAL).
 CC FT SEQUENCE 463 AA; 52942 MW; B48D2FEA3A70A8EA CRC64;
 CC
 CC Query Match 24.0%; Score 54.5; DB 1; Length 463;
 CC Best Local Similarity 28.2%; Pred. No. 16;
 CC Matches 11; Conservative 8; Mismatches 17; Indels 3; Gaps 1;
 CC
 CC 2 KRTSLKNDKFEIKT---DELEIIIGSGSLSTFERLFN 37
 CC Db 213 KEVISFEKERSVDYLMIDIOFGKDSIOEEFFHTFN 251
 CC
 CC RESULT 13
 CC SYFB_PORPU STANDARD; PRT; 720 AA.
 CC ID PS1346;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
 CC tRNA ligase beta chain) (phers).
 CC GN PHER.
 CC OS Porphyra purpurea.
 CC OC Chloroplast.
 CC OX Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 CC NCBI_TaxID=2787;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=AVONPORT;
 CC RA Reith M.E., Munholland J.;
 CC RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
 CC genome.";
 CC RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
 CC diphosphate + L-phenylalanyl-tRNA(Phe).
 CC CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (BY
 CC SIMILARITY).
 CC CC -1- SUBCELLULAR LOCATION: Chloroplast.
 CC CC -1- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
 CC CHAIN FAMILY. SUBFAMILY 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 038804; AAC08232.1;
 CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 CC Chloroplast.
 CC SEQUENCE 720 AA; 82592 MW; 495EA9947117F62A CRC64;

CC Query Match 24.0%; Score 54.5; DB 1; Length 720;
 CC Best Local Similarity 31.5%; Pred. No. 26;
 CC Matches 17; Conservative 7; Mismatches 9; Indels 21; Gaps 2;
 CC
 CC 6 SLKNDKFEIKTDELEIIIGSG-----GSLSTFFLFNRS 39
 CC Db 495 NLKRN-KIETTKLAIILGSLDIRSEMSPAHSLMWYKARKGIENFFKLNKS 547
 CC
 CC RESULT 14
 CC CXB2_MOUSE STANDARD; PRT; 226 AA.
 CC ID CXB2_MOUSE
 CC AC Q00977;
 CC DT 01-APR-1993 (Rel. 25, Created)
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Gap junction beta-2 protein (Connexin 26) (Cx26).
 CC GN GJB2 OR CXN-26.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=92354576; PubMed=1322820;
 CC RA Willecke K., Nicholson B.J., Dahl E., Kozjek G., Hennemann H.;
 CC RT "Molecular cloning of mouse connexins26 and -32: similar genomic
 CC organization but distinct promoter sequences of two gap junction
 CC genes.";
 CC RL Eur. J. Cell Biol. 58:81-89(1992).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=91285228; PubMed=2060697;
 CC RA Nishi M., Kumar N.M., Gilula N.B.;
 CC RT "Developmental regulation of gap junction gene expression during
 CC mouse embryonic development.";
 CC RL Dev. Biol. 146:117-130(1991).
 CC CC -1- FUNCTION: ONE GAP JUNCTION CONSISTS OF A CLUSTER OF CLOSELY PACKED
 CC PAIRS OF TRANSMEMBRANE CHANNELS, THE CONNEXONS, THROUGH WHICH
 CC MATERIALS OF LOW MW DIFFUSE FROM ONE CELL TO A NEIGHBORING CELL.
 CC CC -1- SUBUNIT: A CONNEXON IS COMPOSED OF A HEXAMER OF CONNEXINS.
 CC CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC CC -1- TISSUE SPECIFICITY: LIVER, KIDNEY, INTESTINE, LUNG, SPLEEN,
 CC STOMACH, TESTIS, & BRAIN, BUT NOT HEART & ADULT SKELETAL MUSCLE.
 CC CC -1- SIMILARITY: BELONGS TO THE CONNEXIN FAMILY. BETA-TYPE (GROUP 1)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M81445; AAA37495.1; -
 CC EMBL: M63803; AAA37276.1; -
 CC PIR: C49769; C49769.
 CC DR MGD: MGI:95720; Gjb2.
 CC DR InterPro: IPR000500; Connexin.
 CC DR Pfam: PF00029; connexin; 1.
 CC DR PRINTS: PR00206; CONNEXIN.
 CC DR SMART: SM00037; CNX; 1.
 CC DR PROSITE: PS00407; CONNEXINS_1; 1.
 CC DR PROSITE: PS00408; CONNEXINS_2; 1.
 CC KW Gap junction; Transmembrane.
 CC FT DOMAIN 1 19 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMM 20 40 POTENTIAL.
 CC FT DOMAIN 41 75 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMM 76 96 POTENTIAL.
 CC FT DOMAIN 97 143 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMM 144 164 POTENTIAL.
 CC FT DOMAIN 165 189 EXTRACELLULAR (POTENTIAL).

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:14:14 ; Search time 45.14 Seconds

[without alignments]
176.291 Million cell updates/sec

Title: US-09-833-017-2

Sequence: 1 MKTSLKNDKFKETDELE.....GSLSTFFRLNRSFTQALGK 46

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227	100.0	46	2	099Q15 streptococ
2	220	96.9	46	2	09APK7 streptococ
3	209	92.1	43	2	09APK6 streptococ
4	69	30.4	850	16	097N40 streptococ
5	63.5	28.0	648	2	09L651 streptococ
6	59.5	26.2	275	16	09CPE8 streptococ
7	58	25.6	48	2	033596 streptococ
8	57.5	25.3	1009	16	099WU3 streptococ
9	57.5	25.3	453	3	09C1A3 streptococ
10	57.5	25.3	516	10	09SUK0 streptococ
11	57.5	25.3	516	10	09LX23 streptococ
12	57	25.1	591	2	09AHY8 streptococ
13	56.5	24.9	407	2	09R006 streptococ
14	56.5	24.9	407	2	09L931 streptococ
15	56.5	24.9	408	16	09K17 streptococ
16	56.5	24.9	593	16	092G06 streptococ

17	56	24.7	1250	16	09PEL7 streptococ
18	55.5	24.4	125	12	041368 newcastle d
19	55.5	24.4	125	12	041472 newcastle d
20	55.5	24.4	146	16	092EBO listeria in
21	55.5	24.4	553	12	090339 newcastle d
22	55.5	24.4	553	12	09DLD4 newcastle d
23	55.5	24.4	553	12	09WMH7 newcastle d
24	55.5	24.4	553	12	092150 newcastle d
25	55.5	24.4	553	12	09WLE2 newcastle d
26	55.5	24.4	553	12	09WLE1 newcastle d
27	55.5	24.4	553	12	091HX4 newcastle d
28	55.5	24.4	553	12	091AH8 newcastle d
29	55.5	24.4	837	10	09M819 arabidopsis
30	55.5	24.4	1304	5	09U0H4 plasmodium
31	55	24.2	48	2	033615 streptococ
32	54.5	24.0	230	16	09CGX0 streptococ
33	54.5	24.0	255	2	09AKT3 streptococ
34	54.5	24.0	255	2	09AKN6 streptococ
35	54.5	24.0	398	16	09ALN0 streptococ
36	54.5	24.0	463	16	092H56 streptococ
37	54.5	24.0	819	5	097330 plasmodium
38	54	23.8	48	2	033651 streptococ
39	54	23.8	235	2	09X9J9 vibrio para
40	54	23.8	243	16	097M22 clostridium
41	54	23.8	2353	2	P71401 haemophilus
42	53.5	23.6	125	12	041386 newcastle d
43	53.5	23.6	125	12	041388 newcastle d
44	53.5	23.6	165	1	000392 methanococ
45	53.5	23.6	176	2	09RJ74 streptomyc

ALIGNMENTS

RESULT	ID	SEQUENCE	PRELIMINARY	PRT	46 AA.
099Q15	AC	01-JUN-2001 (TREMblrel. 17, Created)			
099Q15	DT	01-JUN-2001 (TREMblrel. 17, Last sequence update)			
01-DEC-2001 (TREMblrel. 17, Last annotation update)	DE	COMPEXENCE STIMULATING PROTEIN PRECURSOR.			
GN	OS	Streptococcus mutans.			
OC	OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;			
OX	OX	NCBI_TaxID=1309;			
11	RN	SEQUENCE FROM N.A.			
RP	RC	STRAIN-GH14, H7, LT11, NG8, AND UA159;			
RX	RX	MEDLINE-21142515; PubMed-11208787;			
RA	RA	Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitkovitch D.G.;			
RT	RT	"Natural Genetic Transformation of Streptococcus mutans Growing in			
RT	RT	Biofilms."			
RT	RT	J. Bacteriol. 183:897-908(2001).			
DR	DR	EMBL; AF277153; AAK01542.1; -			
DR	DR	EMBL; AF277153; AAK01543.1; -			
DR	DR	EMBL; AF277155; AAK01545.1; -			
DR	DR	EMBL; AF277156; AAK01546.1; -			
DR	DR	EMBL; AF277157; AAK01547.1; -			
DR	DR	InterPro: IPR004288; COMC.			
DR	DR	Plan; PF03047; COMC; 1.			
FT	FT	CHAIN 26			
SO	SO	SEQUENCE 46 AA; 5211 MW; 38FA62B6F78FC3BF CRC64;			

Query Match 100.0%; Score 227; DB 2; Length 46;
Best Local Similarity 100.0%; Pred No. 5.3e-22;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTSLKNDKFKETDELEIIIGSGSLSTFFRLNRSFTQALGK 46
DB 1 MKTSLKNDKFKETDELEIIIGSGSLSTFFRLNRSFTQALGK 46

RESULT 2
 ID 09APK7 PRELIMINARY; PRT; 46 AA.
 AC 09APK7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 GN COMC.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 NX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BM71;
 RX MEDLINE=21142515; PubMed=11208787;
 RA Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitekovich D.G.;
 RT "Natural Genetic Transformation of Streptococcus mutans Growing in
 Biofilms";
 RL J. Bacteriol. 183:897-908(2001).
 DR EMBL: AF277151; AAK01541.1; -
 FT CHAIN 26 46
 SQ SEQUENCE 46 AA; 5195 MW; 38E089DB5B8FC3BF CRC64;

Query Match 96.8%; Score 220; DB 2; Length 46;
 Best Local Similarity 97.8%; Pred. No. 4.2e-21;
 Matches 45; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MKKTLKNDKFKITDELEITIGSGSLTFRLNRSFTQALGK 46
 DB 1 MKKTLKNDKFKITDELEITIGSGSLTFRLNRSFTQALGK 46

RESULT 3
 ID 09APK6 PRELIMINARY; PRT; 43 AA.
 AC 09APK6;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE COMPETENCE STIMULATING PROTEIN.
 GN COMC.
 OS Streptococcus mutans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 NX NCBI_TaxID=1309;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JH1005;
 RX MEDLINE=21142515; PubMed=11208787;
 RA Li Y.H., Lau P.C.Y., Lee J.H., Ellen R.P., Cvitekovich D.G.;
 RT "Natural Genetic Transformation of Streptococcus mutans Growing in
 Biofilms";
 RL J. Bacteriol. 183:897-908(2001).
 DR EMBL: AF277154; AAK01544.1; -
 FT CHAIN 26 43
 SQ SEQUENCE 43 AA; 4927 MW; E6A78FC3BF6156C7 CRC64;

Query Match 92.1%; Score 209; DB 2; Length 43;
 Best Local Similarity 97.7%; Pred. No. 1e-19;
 Matches 42; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKKTLKNDKFKITDELEITIGSGSLTFRLNRSFTQ 43
 DB 1 MKKTLKNDKFKITDELEITIGSGSLTFRLNRSFTQ 43

RESULT 4

097N40
 ID 097N40 PRELIMINARY; PRT; 850 AA.
 AC 097N40;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DE ABC TRANSPORTER, PERMEASE PROTEIN, PUTATIVE.
 GN SP2231.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 NX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouli H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae";
 RL Science 293:498-506(2001)
 DR EMBL: AE007510; AAK06279.1; -
 DR TIGR: SP2231; -
 KW Complete proteome.
 SQ SEQUENCE 850 AA; 97303 MW; 1ADEBD613F06B5115 CRC64;

Query Match 30.4%; Score 69; DB 16; Length 850;
 Best Local Similarity 33.3%; Pred. No. 2.5;
 Matches 19; Conservative 10; Mismatches 16; Indels 12; Gaps 2;

OY 2 KKTLSLKNDKFKITDELEITIGSGSLTFRLNRSFTQALGK 46
 DB 692 KKTFTDNVSEFNLGYTKRKTFTNINVSFGNSQVSFESPTFYRIDTFTTEALQK 748

RESULT 5
 ID 09L651 PRELIMINARY; PRT; 648 AA.
 AC 09L651;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOTHETICAL LACTOCOCCIN 972 IMMUNITY PROTEIN (LACTIS PLASMID PBL1 DNA
 FOR LACTOCOCCIN 972 OPERON).
 GN LCB.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OG Plasmid PBL1.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Lactococcus.
 NX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IPLA 972; PLASMID-PBL1;
 RA Sanchez C., Hernandez de Rojas A., Martinez B., Arqueles M.E.,
 RA Suarez J.E., Rodriguez A., Mayo B.;
 RT "Nucleotide sequence of pBL1, a bacteriocin-producing plasmid from
 RT Lactococcus lactis IPLA 972";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IPLA 972; PLASMID-PBL1;
 RA Suarez J.E.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IPLA 972; PLASMID-PBL1;

RX MEDLINE=20055640; PubMed=10589723;
 RA Martinez B., Fernandez M., Rodriguez A., Saez J.E.;
 RT "Synthesis of lactococcin 972, a bacteriocin produced by *Lactococcus*
 RT *lactis* IPLA 972, depends on the expression of a plasmid-encoded
 RT bicistronic operon."
 RT Microbiology 145:3155-3161(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IPLA 972; PLASMID=PB1;
 RA Martinez B.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF242367; AAF64055.1; -
 DR EMBL: AJ002203; CAC03468.1; -
 KW Plasmid.
 SQ SEQUENCE 648 AA; 74139 MW; 79530E501D6371F7 CRC64;

Query Match 28.0%; Score 63.5; DB 2; Length 648;
 Best Local Similarity 46.2%; Pred. No. 9.6;
 Matches 18; Conservative 8; Mismatches 10; Indels 3; Gaps 2;

QY 5 LSKNDFEIKTDELEIIIGSGSL--STF-FRLFRSF 40
 DB 528 LSVKNRFEIKSQQLFIVTSSIALISSTFLFYNKIV 566

RESULT 6
 ID Q9CPE8 PRELIMINARY; PRT; 275 AA.
 AC Q9CPE8;
 DT 01-JUN-2001 (TREMBLER, 17, Created)
 DT 01-JUN-2001 (TREMBLER, 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLER, 19, Last annotation update)
 DE PROC.
 GN PROC OR PM0095
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of *Pasteurella multocida* PM70."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006044; AAK02179.1; -
 DR InterPro: IPR000304; P5CR.
 DR Pfam: PF01089; P5CR; 1.
 DR PROSITE: PS00521; P5CR; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 275 AA; 29556 MW; 011C7F11AD31A7D7 CRC64;

Query Match 26.2%; Score 59.5; DB 16; Length 275;
 Best Local Similarity 29.6%; Pred. No. 12;
 Matches 16; Conservative 7; Mismatches 16; Indels 15; Gaps 1;

QY 6 LSKNDFEIKTDELEIIIGSGSLSTFRLFRNSFTQAL 44
 DB 141 LSKNDFEIKTDELEIIIGSGSLSTFRLFRNSFTQAL 194

RESULT 7
 ID Q33596 PRELIMINARY; PRT; 48 AA.
 AC Q33596;
 DT 01-JAN-1998 (TREMBLER, 05, Created)
 DT 01-JAN-1998 (TREMBLER, 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLER, 19, Last annotation update)
 DE COMPETENCE STIMULATING PEPTIDE PRECURSOR.
 GN COMC.
 SQ Streptococcus anginosus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus;
 OX NCBI_TaxID=1328;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIC 10713;
 RX MEDLINE=98012953; PubMed=9352904;
 RA Havarstein L.S., Hakenbeck R., Gaustad P.;
 RT "Natural competence in the genus *Streptococcus*: Evidence that
 RT streptococci can change phenotype by interspecies recombinational
 RT exchanges."
 RL J. Bacteriol. 179:6589-6594(1997).
 DR EMBL: AJ000864; CAA04341.1; -
 DR InterPro: IPR004288; COMC.
 DR Pfam: PF03047; COMC; 1.
 KW Signal.
 FT SIGNAL.
 FT CHAIN 33 48
 SQ SEQUENCE 48 AA; 5621 MW; EB6DF7AEDE4E94DB CRC64;

Query Match 25.6%; Score 58; DB 2; Length 48;
 Best Local Similarity 30.2%; Pred. No. 2.8;
 Matches 16; Conservative 9; Mismatches 16; Indels 12; Gaps 2;

QY 1 MKRTLSLRN-----DKRKITDELEIIIGSGSLSTFRLFRNSFTQALG 46
 DB 1 MKRLFAKKEVYKVEKRELNDQLKIIIGSDSRIMGF----DPSKLEGR 48

RESULT 8
 ID Q99WU3 PRELIMINARY; PRT; 1009 AA.
 AC Q99WU3;
 DT 01-JUN-2001 (TREMBLER, 17, Created)
 DT 01-JUN-2001 (TREMBLER, 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLER, 19, Last annotation update)
 DE SA0272 PROTEIN (HYPOTHETICAL PROTEIN SAV0283).
 GN SA0272 OR SAV0283.
 OS *Staphylococcus aureus* (strain N315), and
 OS *Staphylococcus aureus* (strain Mu50).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=158879, 158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-O., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani M., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuuchi J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogatawa N., Hayashi H., Hiratsuka K.;
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus*
 RT *aureus*."
 RL Lancet 357:1225-1240(2001).
 DR EMBL: AP003130; BAB41496.1; -
 DR EMBL: AP003358; BAB56445.1; -
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 1009 AA; 114781 MW; 16AD06D170CA31FE CRC64;

Query Match 25.6%; Score 58; DB 16; Length 1009;
 Best Local Similarity 34.3%; Pred. No. 79;
 Matches 12; Conservative 8; Mismatches 15; Indels 0; Gaps 0;

QY 9 NDKRKITDELEIIIGSGSLSTFRLFRNSFTQAL 43
 DB 202 NDKRKITDELEIIIGSGSLSTFRLFRNSFTQAL 236

RESULT	9			
09C1A3				
ID	09C1A3	PRELIMINARY;	PRT:	453 AA.
AC	09C1A3:			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
RT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)			
DE	SULFATE TRANSPORTER SUL2-LA (FRAGMENT).			
OS	Saccharomyces pastorianus (yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=27292;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	James A.B., Slaughter C., Meaden P.G.;			
RT	"Characterization of partial coding sulfate transporter sequences from			
RT	Saccharomyces pastorianus and Saccharomyces bayanus.";			
RL	Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL: AF364410; AAK32879.1; -			
DR	Interpro: IPR001902; Sulfate_transp.			
DR	Pfam: pf00916; Sulfate_transp; 1.			
FT	NON_TER	1		
FT	NON_TER	453		
SO	SEQUENCE	453 AA;	49519 MW;	137117AD012BD31F CRC64;

Query Match	25.38;	Score 57.5;	DB 3;	Length 453;
Best Local Similarity	36.7%;	Pred. No. 38;		
Matches	18;	Conservative	6;	Mismatches 18; Indels 7; Gaps 2

```
QY      1 MKTTLKNDKFEKIDLEIIIGSGSLSTFRL-----FNRSFTA 43
      : | : | : | : | : | : | : | : | : | : | : | : |
Db      227 ISKSGRVNDYKVPDQEL-IALGVSNLLGTFNNAYPATGSFSRSALKA 274
```

RESULT	10			
09SUK0	ID	09SUK0	PRELIMINARY;	PRT; 516 AA.
AC	AC	09SUK0		
DT	01-MAY-2000	(TREMBLrel. 13, Created)		
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
DE	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).			
GN	ACGN.			
OS	Arabidopsis thaliana (mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99364543; PubMed=10437832;			
RA	Wendt U.K., Hauschild R., Lange C., Pietersma M., Wenderoth I.,			
RA	von Schewen A.;			
RT	"Evidence for functional convergence of redox regulation in G6PDH			
RT	isoforms of cyanobacteria and higher plants.";			
RL	Plant Mol. Biol. 40:487-494(1993).			
CC	-1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCONO-			
CC	DELTA-LACTONE 6-PHOSPHATE + NADPH.			
CC	-1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.			
CC	-1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE			
CC	FAMILY			
EMBL	AF010970; CAB52674.1; -.			
DR	HSSP; P11411; IDPG.			
DR	InterPro; IPR001282; G6PD.			
DR	Pfam; PF00479; G6PD_1.			
DR	Pfam; PF02781; G6PD_C; 1.			
DR	PRINTS; PR00079; G6PDHGNASE.			
DR	ProDom; PD001129; G6PD.1.			
DR	PROSITE; PS00069; G6P_DEHYDROGENASE.1			
KW	Glucose metabolism; NADP; Oxidoreductase.			
QO	SOURCE 516 AA; 59171 MW; A/623780F053CA1C CRC64;			

Query Match	25.38;	Score 57.5;	DB 10;	Length 516;
Best Local Similarity	41.58;	Pred. No. 44;		
Matches 22; Conservative	7;	Mismatches 11;	Indels 13;	Gaps 5

```
QY      1 MKKTLSLKND--FKE---IKTDELEIII-GGSGSLS--TF---ERLFNRSE 400
        | : | : ||| | | : | : | : | : | : | : | : | : | : |
Db      8 MEKRSTLKNDSFVKEYNPIYETGSISIVLGASGLAKKTPEPALNLFHQGF 600
```

RESULT	11
Q9LK23	
TD	00W33
DDP	THRUWAY
SEE	E1C 11

DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (EC 1.1.1.49) (G6PD).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CX NCBI_taxonomy:3702;

RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA,
RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databass.

RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;

RT Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P
RT TAC and BAC clones." ;
RL DNA Res. 7:217-221(2000).
CC -1- CATALYTIC ACTIVITY: D-GLUCOSE 6-PHOSPHATE + NADP(+) = D-GLUCO-6-
CC DELTA-LACTONE 6-PHOSPHATE + NADPH.
CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
CC FAMILY.
DR EMBL: AP000381; BAB02125.1; -.
DR HSSP: E1411; IDPG.
DR InterPro: IPR001282; G6PD.
DR Pfam: PF004479; G6PD; 1.
DR DR Pfam: PF02783; G6PD_C.1.
DR PRINTS: PR00079; G6PDDKDNASE.
DR PRODOM: PD001129; G6PD; 1.
DR PROSITE: PS00069; G6P_DEHYDROGENASE; 1.
RW Glucose metabolism; NADP; Oxidoreductase.
SQ SEQUENCE 516 AA; 5157 MW; 407EA2590EFC20E2 CRC64;

Query Match	25.38;	Score 57.5;	DB 10;	Length 516;
Best Local Similarity	41.58;	Pred. No. 44;		
Matches 22; Conservative	7;	Mismatches 11;	Indels 13;	Gaps 5;

```

Qy      1 MKKTLSLKND--EKE----IKTDELEIIT--GGSSLS---TF---EFLFNRSF 400
      1:1: :|||  || :||: |||: || :||:|
Db      8 MEKRS TLKNDSEFVKEYNPVTETEGSLIIVLGASDLAKKTFPALENLFHQGF 600

```

RESULT

ID	Q9AHY9	PRELIMINARY;	PRT;	591 AA
----	--------	--------------	------	--------

OS Photorhabdus luminescens (Xenorhabdus luminescens)
GN
DE VFPA. (FRAGMENT).
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update).
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update).
DT 01-JUN-2001 (TREMBLrel. 17, Created)

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Photobacterium;
 OX NCBI_TaxID=29488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NC19, ATCC29304;
 RX MEDLINE-21225535; PubMed=11325940;
 RA Cliche T.A., Brinton S.B., Horswill A.R., Engen J.C.;
 RT "A Phosphopantetheinyl Transferase Homolog Is Essential for
 Photobacterium luminescens to Support Growth and Reproduction of the
 Endosymbiotic Nematode Heterorhabditis bacteriophora.";
 RL J. Bacteriol. 183:3117-3126(2001).
 DE EMBL; AF288085; AAK16098.1;
 GN HSP; P39435; LKAS.
 OS Interpro: IPR000794; Ketoacyl-synt.
 DR Pfam: PF00109; ketoacyl-synt_C.1
 DR PROSITE: PS00606; B_KETOACYL_SINTHASE; UNKNOWN_1.
 FT NON_TER 591
 SQ SEQUENCE 591 AA; 63878 MW; AFI451C6DD500891 CRC64;

Query Match 25.1%; Score 57; DB 2; Length 591;
 Best Local Similarity 34.3%; Pred. No. 59;
 Matches 12; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

OY 12 KEIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 149 QRVNRDRIGVILGGGNGMTLFLSLGRQOTPYLRK 183

RESULT 13
 O9R006 PRELIMINARY; PRT; 407 AA.
 AC O9R006; (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE BUSA.
 GN BUSA.
 OS Lactococcus lactis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX NCBI_TaxID=1358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NDO 763;
 RX MEDLINE-99445450; PubMed=10515910;
 RA Obis D., Guillot A., Girpon J.C., Renault P., Bolotin A., Mistou M.Y.;
 RT "Genetic and biochemical characterization of a high-affinity betaine
 uptake system (BUSA) in Lactococcus lactis reveals a new functional
 RT organization within bacterial ABC transporters.";
 RL J. Bacteriol. 181:6238-6246(1999).
 DE EMBL; AF139575; AAF04258.1;
 GN EMBL; AF139575; AAF04258.1;
 OS Interpro: IPR003593; AAA.
 DR Interpro: IPR003439; ABC_transportr.
 DR Interpro: IPR001687; ATP_GTP_A.
 DR Interpro: IPR000644; CBS.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00116; CBS; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 407 AA; 45696 MW; D5DFBC0F8764B3A9 CRC64;

Query Match 24.9%; Score 56.5; DB 2; Length 407;
 Best Local Similarity 45.7%; Pred. No. 46;
 Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 13 EIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 149 QRVNRDRIGVILGGGNGMTLFLSLGRQOTPYLRK 183

DB 50 EINEGEIIVIMGLSGSKSTLLRLNRIEPTSGK 84

RESULT 14
 O9L931 PRELIMINARY; PRT; 407 AA.
 AC O9L931;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE OPVUA.
 GN OPVUA.
 OS Lactococcus lactis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX NCBI_TaxID=1358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MG1363;
 RX MEDLINE-20393175; PubMed=10939245;
 RA Bouvier J., Bordes P., Romeo Y., Fourcans A., Bouvier I.,
 RA Gutierrez C.;
 RT "Characterization of OpvA, a glycine-betaine uptake system of
 Lactococcus lactis.";
 RL J. Mol. Microbiol. Biotechnol. 2:199-205(2000).
 DE EMBL; AF184955; AAF0473.1;
 GN EMBL; AF184955; AAF0473.1;
 OS Interpro: IPR003593; AAA.
 DR Interpro: IPR003439; ABC_transportr.
 DR Interpro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00116; CBS; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 407 AA; 45653 MW; 656D9097AE58407F CRC64;

Query Match 24.9%; Score 56.5; DB 2; Length 407;
 Best Local Similarity 45.7%; Pred. No. 46;
 Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

OY 13 EIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 50 EINEGEIIVIMGLSGSKSTLLRLNRIEPTSGK 84

RESULT 15
 O9K1F7 PRELIMINARY; PRT; 408 AA.
 AC O9K1F7;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE OPVUA (BETAINE ABC TRANSPORTER ATP BINDING PROTEIN).
 GN OPVUA OR BUSA.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis)
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-1L403;
 RX MEDLINE-20318987; PubMed=10860977;
 RA van der Heide T., Poolman B.;
 RT "Osmoregulated ABC-transport system of Lactococcus lactis senses water
 RT stress via changes in the physical state of the membrane.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:7102-7106(2000).
 DE EMBL; AF184955; AAF0473.1;
 GN EMBL; AF184955; AAF0473.1;
 OS Interpro: IPR003593; AAA.
 DR Interpro: IPR003439; ABC_transportr.
 DR Interpro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00116; CBS; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 407 AA; 45653 MW; 656D9097AE58407F CRC64;

OY 13 EIKTDELEIIGSGSLSTFFRLFNRSFTQALGK 46
 DB 50 EINEGEIIVIMGLSGSKSTLLRLNRIEPTSGK 84

RA Weissenbach J., Ehrlich S.D., Sorokin A.
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RL *lactis* ssp. *lactis* IL1403.";
 CC Genome Res. 11:731-753(2001).
 -) SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS).
 DR EMBL: AF234619; AAF37878.1; -;
 DR EMBL: AE006375; AK05550.1; -;
 DR InterPro; IPR003593; AAA.
 DR InterPro; IPR003439; ABC_transport.
 DR InterPro; IPR001687; ATP_GTP_A.
 DR InterPro; IPR000644; CBS.
 DR Pfam; PF00005; ABC_tran; 1.
 DR Pfam; PF00571; CBS; 2.
 DR SMART: SM00382; AAA; 1.
 DR SMART: SM00116; CBS; 2.
 DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 DR ATP-binding; Complete proteome; transport.
 KW ATP-binding; 408 AA; 45750 MW; EBBAB87BE563B88 CRC64;
 SQ SEQUENCE

Query Match 24.9%; Score 56.5; DB 16; Length 408;
 Best Local Similarity 45.7%; Pred. No. 46;
 Matches 16; Conservative 3; Mismatches 15; Indels 1; Gaps 1;

QY 13 EIKTDELEITIGSGS-LSTFRLFNRSFTQALGK 46
 DB 50 EINEGEIFVIMGLSGSKSTLLRLNRLLEPTSGK 84

Search completed: July 30, 2002, 15:20:30
 Job time: 376 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:14:11 ; Search time 53.42 Seconds

(without alignments)
916.951 Million cell updates/sec

Title: US-09-833-017-4

Perfect score: 2173

Sequence: 1 MNEALMILSLGLTYLVLF.....KTSLSQSNHHLFKOLLILK 441

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq.032802:*

- 1: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-geneeq/geneeqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	752	34.6	446	20	AAV14065
2	752	34.6	446	20	AAV03650
3	752	34.6	446	20	AAW89437
4	752	34.6	446	20	AAW89434
5	398	18.3	175	19	AAH85911
6	378.5	17.4	432	22	AAH81843
7	258	11.9	108	19	AAW38531
8	192	8.7	106	19	AAW38530
9	189	8.7	106	18	AAV11170
10	151	6.9	543	22	AAW98391
11	145.5	6.7	442	20	AAW83369

12	145.5	6.7	442	21	AAV57684
13	143	6.6	861	19	AAW69977
14	143	6.6	909	22	AAU37397
15	142.5	6.6	363	22	AAU37208
16	141.5	6.5	363	19	AAW75788
17	139	6.4	605	22	ABW52567
18	135.5	6.2	478	21	AAV95040
19	135.5	6.2	1120	22	AAU15085
20	134	6.2	365	19	AAW98798
21	134	6.2	565	19	AAW71554
22	134	6.2	298	22	AAH6327
23	133	6.1	298	22	AAH62102
24	125.5	5.8	4134	20	AAH31946
25	124.5	5.7	952	22	ABG29521
26	124	5.7	451	22	AAH81558
27	122	5.6	436	22	AAU03143
28	119	5.5	600	17	AAH95269
29	119	5.5	618	20	AAV06667
30	117	5.4	618	21	AAH28685
31	116	5.3	1477	20	AAV19981
32	116	5.3	1494	20	AAV19980
33	115.5	5.3	1079	22	ABH69644
34	115	5.3	850	19	AAW57445
35	115	5.3	1167	22	ABH81844
36	114.5	5.3	904	20	AAW6351
37	114	5.2	457	20	AAV05736
38	114	5.2	543	22	ABH47270
39	113.5	5.2	443	21	AAH81506
40	113.5	5.2	451	20	AAV28601
41	113	5.2	1865	17	AAW03515
42	112.5	5.2	471	20	AAV20061
43	112.5	5.2	490	20	AAV20060
44	112.5	5.2	1346	21	AAH18236
45	112	5.2	794	22	AAH00426

ALIGNMENTS

RESULT 1

AAV14065 standard; Protein: 446 AA.

AAV14065;

16-JUL-1999 (first entry)

S. pneumoniae Ornithine carbamoyltransferase.

Ornithine carbamoyltransferase: argF/arcB family polypeptide; diagnosis; infection; therapy; immune response; otitis media; conjunctivitis; pneumonia; bacteraemia; sinusitis; pleural empyema; endocarditis; meningitis; matrix protein adhesion.

Streptococcus pneumoniae.

EP913476-A2.

06-MAY-1999.

22-OCT-1998; 98EP-0203571.

30-OCT-1997; 97US-0961536.

(SMK) SMITHKLINE BEECHAM CORP.

Brown JR, Zalacain M;

WPI: 1999-256630/22.

N-PSDB: AAX57966.

New Streptococcus pneumoniae ornithine carbamoyltransferase (argF/arcB) polypeptide and polynucleotide, us

Streptococcus pneu
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Escherichia coli p
Candida albicans p
Protein encoded by
H. pylori GPO 128
Helicobacter poly
H. pylori HPS024 p
S. epidermidis ope
Plasmodium falcipa
Novel human diagno
S. epidermidis ope
Streptococcus pyog
Nisin nist gene pr
Nisin C of lactoba
Staphylococcus war
B. burgdorferi ant
B. burgdorferi ant
Drosophila melanog
A. thermophilum th
Human DNA toll-11
Staphylococcus aure
Enterococcus faeca
Streptococcus pneu
Histidine kinase p
Human DOKK180 prot
B. burgdorferi ant
Plasmodium falcipa
P. falciparum telo

This sequence is the *Streptococcus pneumoniae* ornithine carbamoyltransferase protein is a (argF/argB) family polypeptide. (I) and polymucleotides (II) encoding it are useful for diagnosing diseases due to an infection of an organism with the ornithine carbamoyltransferase gene by determining the nucleic acid sequence encoding (I), and/or analysing for the presence or amount of (I). They can diagnose the stage and type of infection. (I) is also useful for screening for compounds which affect activity of the protein by measuring the binding to (I) and observing the stimulation or inhibition of the polypeptide function. These can be used in treatment to inhibit or enhance ornithine carbamoyltransferase activity, in addition to direct administration of ornithine carbamoyltransferase polypeptides to treat conditions associated with a lack of ornithine carbamoyltransferase polypeptide, or direct administration of antisense sequences to prevent expression. (I) and antibodies against it induce an immune response to immunise and prevent disease. Anti-(I) antibodies induced by the polypeptide are also useful for isolating clones expressing (I), or for purifying the polypeptide by affinity chromatography. Diseases diagnosed, prevented or treated include cystitis media, conjunctivitis, pneumonia, bacteraemia, sinusitis, pleural effusion, endocarditis and especially meningitis. (I), (II) and their (ant)agonists can prevent adhesion of bacteria to matrix proteins, and are useful for use on wounds and body implants to prevent bacterial infection.

Note: the specification states that this sequence is identical to that shown in *AV144067*.

Query Match	34.6%;	Score 752;	DB 20;	Length 446;
Best Local Similarity	39.2%;	Pred. No. 3e-58;		
Matches 172; Conservative	86;	Mismatches 169;	Indels 12;	Gaps 5

```

OY 12 LLLFLVLA--FLFLFSPKSVNVAUSKKEELFLPSINSFILMTAVT-----MNVNLAFLPYPAE 64
Db 6 lllvalvngklavifkfkngvigtldtctrfkafllkllgllftctfqflavskylsyfie 65
OY 65 PLVFLALSIYLNRONSLSLNIFFYGLLPVASSDLFRRAIIFFILDOGT-OGIV--MGSSIIT 121
Db 66 plfglfgtstlllltqlpkklllfgllfpmllvelfygvysvlpflg9gldvgdgnbifl 125
OY 122 TYMIEFPGALISYLFSLFSEVNDIGRKDSLTKMVKYKRILPBNITMLLYLLIOVLVIE 161
Db 126 lllmflvcfiyI--vflkwladydctrlrfelbtqfqsstklnwamgasyllvmgslylie 163
OY 182 SYNVIPTLKEFEKRVYIYLLFLPILSLFSLTOYTKORVONETIMAEKOAIRNITQYSOOIE 241
Db 184 yegllgetvtrhhlllvuyllftmggljkkldtylkkelgelneqetllyrdmerysrhie 243
OY 242 SLKYDINSFRHDYINILITSRLGSLIENKDLASIEKTIYHQILEKGTGHQLODPYXNIGHLANI 301
Db 244 elykeitsrtfndytnllstlrllgieedmeqelkydsvllrdsqklqdnkxydlgrvni 303
OY 302 ONDAVKGILSLKILEAONKRIANVNVSSKIDLPKEMELDLFTTILSICDNAPAFESTL 361
Db 304 rdlralkslldgkflkateknlvfnoveypraeiayegmsllldftllvlllclnahaivesaas 363
OY 362 NPEIOLAEFFKKNGSIVETIGNSTRKEOIODVSKIFKEKNTSGKNSGNGIGLAVNHIIEHPY 421
Db 364 qphvsiatllknagqetflensikeegldlseifsfgaaskeorgevgylyvmwkliveshp 423
OY 422 KTSIOTSNHNLHFKOLLII 440
Db 424 ntnlnttcqgvtrgvltv 442

```

RESULT 2

QY 65 PLYFIALSIYLNKQNSLSLNIFFGLLPVASSDLERRAIIFFILBDGT-QGIV--MGSSILIT 121

Dib

OY 122 TYMIEFAGIALSYFLSVFNVDIGRLKDSLTFRKVRKRLPMNITMLLYLLIOVLVIE 181
 CC the binding, or stimulation or inhibition of RR protein activity.
 CC Treatment of diseases related to protein activity can be treated by
 CC administering antagonists to prevent expression, and RR proteins to
 CC enhance expression levels of the protein. RR proteins and polynucleotides
 CC can be administered as a vaccine to protect against disease, and can
 CC prevent adhesion of bacteria to matrix proteins, and are useful for use
 CC on wounds and body implants to prevent bacterial infection. Antibodies
 CC are useful for treating bacterial infections, and can isolate or identify
 CC clones expressing RR proteins, and can purify the proteins by affinity
 CC chromatography. Diseases diagnosed, prevented or treated include: otitis
 CC media; conjunctivitis; pneumonia; bacteremia; sinusitis; pleural empyema;
 CC endocarditis and particularly meningitis, and infections caused by
 CC Streptococcus pneumoniae.
 CC XX
 SQ Sequence 446 AA:
 OY 362 NPEIOLAFKKNGSIVFIIONSTREKOIDVSKIFKENTSTKSGNIGLAKVNHILEHYP 421
 Db 364 qphvsiaflknagqetfllensikeegidseifsgasskgeergvgylyvmkiveshp 423
 OY 422 KTSIOTSNNHHFLKOLLIT 440
 Db 424 ntlnltcgnqvtgvtlv 442
 RESULT 3
 AAM89437 standard; Protein: 446 AA.
 ID AAM89437:
 AC AAM89437:
 XX 22-MAR-1999 (first entry)
 DT 22-MAR-1999 (first entry)
 DE Streptococcus pneumoniae histidine kinase.
 KW Streptococcus pneumoniae; response regulator; histidine kinase;
 KW infection; antibacterial; otitis media; conjunctivitis; pneumonia;
 KW bacteremia; sinusitis; pleural empyema; endocarditis; meningitis.
 OS Streptococcus pneumoniae.
 XX
 PN EP885961-A2.
 XX 23-DEC-1998.
 PD 09-JUN-1998; 98EP-0304562.
 PF 13-JUN-1997; 97US-0874138.
 PR (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Wallis NG:
 XX WPI, 1999-037018/04.
 DR N-PSDB; AAV82066.
 XX
 PT New response regulator polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of
 PT Streptococcus pneumoniae infections
 XX
 PS Disclosure: Page 16; 40pp; English.
 CC The present sequence represents histidine kinase from Streptococcus
 CC pneumoniae, which is cognate of the response regulator (RR) of the
 CC present invention. RR proteins and polynucleotides are useful as
 CC research reagents and as materials for discovery of treatments of, and
 CC for diagnosing diseases, particularly for diagnosing the stage and type
 CC of an infection using PCR. Diseases related to expression or activity of
 CC RR proteins can be diagnosed by determining the nucleotide sequence
 CC encoding RR, or by analysing for the presence or amount of RR proteins.
 CC RR proteins can be used to screen for agonists and antagonists

CC (antibacterial compounds) which interact with RR proteins by observing
 CC the binding, or stimulation or inhibition of RR protein activity.
 CC Treatment of diseases related to protein activity can be treated by
 CC administering antagonists to prevent expression, and RR proteins to
 CC enhance expression levels of the protein. RR proteins and polynucleotides
 CC can be administered as a vaccine to protect against disease, and can
 CC prevent adhesion of bacteria to matrix proteins, and are useful for use
 CC on wounds and body implants to prevent bacterial infection. Antibodies
 CC are useful for treating bacterial infections, and can isolate or identify
 CC clones expressing RR proteins, and can purify the proteins by affinity
 CC chromatography. Diseases diagnosed, prevented or treated include: otitis
 CC media; conjunctivitis; pneumonia; bacteremia; sinusitis; pleural empyema;
 CC endocarditis and particularly meningitis, and infections caused by
 CC Streptococcus pneumoniae.
 CC XX
 SQ Sequence 446 AA:
 Query Match 34.6%; Score 752; DB 20; Length 446;
 Best Local Similarity 39.2%; Pred. No. 3e-58;
 Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;
 OY 12 LTYLYNVL--FLFLFSSKVSNTLSKKELTLPISNFTIMIAVW----WNNVNLFFPAE 64
 Db 6 llyalvnglkivlftfkvngqgltdtrfkafklitllfttqflavskyslyfie 65
 OY 65 PLVFIALSTYLNKONSLSNIEFGLLPVASSDFFRAIIFELDGT--QGIV--MGSSIIT 121
 Db 66 plfgjsgisflllrglpkkliffyglfpmllvelfyrgvsyfvlpfjgglvdgggnpfl 125
 OY 122 TYMIEFAGIALSYFLSVFNVDIGRLKDSLTFRKVRKRLPMNITMLLYLLIOVLVIE 181
 Db 126 lmlfvcfvl--vflkwldydftrlrrelfdtqfksltknwamgasylyvmgslyle 183
 OY 182 SYNVIPFLKFRKRVIVYVILFLLISLSOYTKOKVONEMAKOIRNITOYSOQIE 241
 Db 184 yegqisltvtrhllvlyllffmgikldtylkeklgeelngeltlryimemysrhe 243
 OY 242 SLVKDISFRHDYLNITLSRLGIENKDLASIEKTYHOLEKTHODPTRYNIGHLANI 301
 Db 244 elykeisfrhdynltlstrlgdeedmegkelydsvlrdsqklgdnkydlgrlvni 303
 OY 302 QNDVAKGILSAKTLEAONKRIAVNVESSKIQLPMEMLDPTITLSILCDNAIEAFESL 361
 Db 304 rdtalksllagkfkikareknlvfvevpeelqvegmslldflvslldcnaievsaas 363
 OY 422 KTSIOTSNNHHFLKOLLIT 440
 Db 424 ntlnltcgnqvtgvtlv 442
 RESULT 4
 AAM89434 standard; Protein: 446 AA.
 ID AAM89434:
 AC AAM89434:
 XX 22-MAR-1999 (first entry)
 DT 22-MAR-1999 (first entry)
 DE Streptococcus pneumoniae histidine kinase.
 KW Streptococcus pneumoniae; response regulator; histidine kinase;
 KW infection; antibacterial; otitis media; conjunctivitis; pneumonia;
 KW bacteremia; sinusitis; pleural empyema; endocarditis; meningitis.
 OS Streptococcus pneumoniae.
 XX
 PN EP885963-A2.

PD	23-DEC-1998.	
XX		
PF	17-JUN-1998;	98EP-0304797.
XX		
PR	20-JUN-1997;	97US-0879528.
XX		
PA	(SMIK) SMITHKLINE BEECHAM CORP.	
PA	(SMIK) SMITHKLINE BEECHAM PLC.	
XX		
XX		
P1	Mallis NW;	
XX		
DR	WPI: 1999-037058/04.	
DR	N-PSDB: AAN82059.	
XX		
PT	New response regulator polypeptide and polynucleotide - useful as	
PT	diagnostic reagents and for prevention and treatment of	
PT	Streptococcus pneumoniae infections	
XX		
PS	Disclosure: Page 15; 37pp; English.	
XX		
CC	The present sequence represents histidine kinase from Streptococcus	
CC	pneumoniae, which is cognate of the response regulator (RR) of the	
CC	present invention. RR proteins and polynucleotides are useful as research	
CC	reagents and materials for discovery of treatments of and diagnostics	
CC	for disease, and particularly for diagnosing the stage and type of an	
CC	infection by PCR. Diseases related to expression or activity of RR can	
CC	be diagnosed by determining the nucleotide sequence encoding RR, or by	
CC	analysing for the presence or amount of RR protein. RR proteins can be	
CC	used to screen for agonists and antagonists which interact with RR	
CC	protein by observing the binding, or stimulation or inhibition of RR	
CC	protein activity. The agonist can be used in treatment to inhibit RR	
CC	activity, and RR proteins can be administered to treat conditions	
CC	associated with a lack of RR protein. Diseases can be prevented by	
CC	inoculating with RR protein, or a vector which expresses RR protein. RR	
CC	proteins and polynucleotides can prevent adhesion of bacteria to matrix	
CC	proteins, and are useful for use on wounds and body implants to prevent	
CC	bacterial infection. RR antibodies are useful for treating bacterial	
CC	infections, and can isolate or identify clones expressing RR proteins,	
CC	and can purify the proteins by affinity chromatography. Diseases	
CC	diagnosed, prevented or treated include: otitis media; conjunctivitis;	
CC	pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis and	
CC	particularly meningitis, and infections caused by Streptococcus	
XX		
XX		
Sequence	446 AA;	
50		

	Query Match:	34.6%;	Score 752;	DB 20;	Length 446;
	Best Local Similarity:	39.58%;	Pred. No. 3e-58;		
	Matches 172;	Conservative	86;	Mismatches 169;	Indels 12; Gaps 5
OY	12	LATYLVVL--ELLEFSKVSNTLSKEKELTLESISNELIMIAVN-----MVNMLEYPAE	64		
Db	6	lllyavtngljkivillffkvngigltfdkifkaflllkllgllflftfgflaskyslfyle	65		
OY	65	PVFEALSTVLNQNSLCSLNIFEGCLLPVASSDPRRAIIEFLIOGT--OGIV--NGSSITP	121		
Db	66	plfvgislllllgldrkklilfyglfmlveelfryvsyrvlbfllqagvldgggnpifl	125		
OY	122	TWMEFAGIALSYLFSEVENVDIGRLKSLRKMKYKKRLDEMNITMLLYLLIQVLVIYE	181		
Db	126	llimfcvcfiwl-vfkikldgydfrtlrrefldtgfskcltkinamgayzlvmqslayle	183		
OY	182	SYNVIPPLKRKRKVNVILVYLILFFLILSPLSOYTKRKNQNEIMAKOEAOINRIHQYSQIE	241		
Db	184	yegdigvtcrvnllivfyllffimggijkdktylkekleeeenqbcrltyrdomeysrhie	243		
OY	242	SLXKDINSFRHDIPLNLITSRLGIEENKLASIEKYIHOILEKTGHODLRFYNNIGHLANI	301		
Db	244	eilyekstsrfhzytnlltsrlrgieedmeqikeileylvdlsdqskldqnkydlgrlvnl	303		
OY	302	QNDAVKGLSAKLLEONKKRIAVNVBYSKIDLPHEMLLDPTTLISILCONATAAESLT	361		

Db 304 rdaLksllagkfikareknivfwevreeqvegmssldfltlvislclodalvsaas 363

Oy 362 NPEtOAFKFKKGSVYFIIIONSTKEKODVSKFKFKNYSTKSNNGIGLAKVNIILEHP 421

Db 364 qpnvstaflkngaqefliemsikeegldtseltsgaaskgeevyglvlymkvlveshp 423

Oy 422 KTSLSOTSNHHNHLKOLLII 440

Db 424 ptnlntcagvdfvrgvltv 442

RESULT	5	
ID	AAV85911	
XX	AAV85911 standard; Protein; 175 AA.	
AC		
XX	AAV85911;	
DT	10-APR-2000 (first entry)	
XX		
DE	S. pneumoniae derived protein #120.	
XX		
RW	Treatment; prevention; disease; diagnosis; gene therapy; screening;	
KW	bacterial; antimicrobial; antibiotic; pathogenesis; infection.	
XX		
OS	Streptococcus pneumoniae.	
XX		
PN	WO9806734-A1.	
XX		
PD	19-FEB-1998.	
XX		
PF	15-AUG-1997; 97WO-US14436.	
XX		
PR	16-AUG-1996; 96US-0024022.	
XX		
PA	(SMIK) SMITHKLINE BEECHAM CORP.	
XX		
PI	Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;	
PI	Stodola RK;	
XX		
DR	WPI: 1998-159452/14.	
XX		
DR	N-PSDB: AA296278.	
XX		
PT	Streptococcus pneumoniae proteins and related DNA - useful for	
PT	screening compounds for antibacterial activity	
XX		
SS	Claim 5; Page 404-405; 640pp; English.	

CC This invention describes novel isolated *Streptococcus pneumoniae*
CC polynucleotides (see AA296173-296494) and their encoded proteins (see
CC AY85792-968182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The also
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.

XX
XX Sequence 175 AA:

Query Match 18.3%; Score 398; DB 19; Length 175;
Best Local Similarity 44.4%; Pred. No. 2,1e-27;
Matches 76; Conservative 40; Mismatches 55; Indels 0; Gaps 0;

QY 270 LASTEKTTHDLEKKTGHOADTRRYNGHANTQNDVAVGLTSAAKLEAGKRIAAVVEVS 329
Db 1 meqikekyelovsltdssqklyqnhkydgltyivtrdralsltaagfllakareknlvinyevp 60

Novel *Streptococcus pneumoniae* proteins and related DNA - useful for diagnosing anti-microbial agents for treatment of bacterial

DR N-PSDB; AAX30770

DR N-PSDB; AAX30770.
XX

CC	AAAG98391 standard; Protein; 543 AA.
AC	AAAG98391;
AD	21-SEP-2001 (first entry)
AE	Escherichia coli protein sequence SEQ ID NO:439.
AF	Escherichia coli; identification; proliferation; microorganism;
AG	antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
AH	bacterial growth inhibition.
AI	Escherichia coli.
AJ	MO200148209-A2.
AK	05-JUL-2001.
AL	19-DEC-2000; 2000WO-US34419.
AM	23-DEC-1999; 99US-0173005.
AN	(ELIT-) ELITRA PHARM INC.
AO	Forsyth RA, Ohlsen KL, Zyskind JW;
AP	WPI; 2001-457376/49.
AQ	N-PSDB; AAH81447.
AR	Novel nucleic acids encoding proteins required for Escherichia coli
AS	proliferation, useful for screening for antimicrobial agents -
AT	Claim 19; Page 558-559; 596pp; English.
CA	The present invention describes a purified or isolated nucleic acid
CB	sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC	given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CD	microorganism is capable of inhibiting proliferation of a microorganism.
CE	(I) have antibacterial and antibiotic activities, and can be used in
CF	gene therapy. Expression of (I) in a microorganism inhibits proliferation

```
Query Match      6.9%   Score 151; DB 22; Length 543;  
Best Local Similarity    23.4%; Pred. No. 8,2e-05;  
Matches     89; Conservative     68; Mismatches 143; Indels    80; Gaps    18;
```

```
QY      85 IFVGLPLVASDILF--RRAIIFFILDGTOGIYMSSITTYMIEFAIALSYL-----F 136  
       |||:|||||::|||::|::|::|::|::|::|::|::|::|:  
Db    208 IIfIepYeIslfIdgrqamIgslggyvaavaddgey--clindaagellnlytsqddeK 265  
              |  
QY    137 LSVEN----VDIGR-LKDSELR---MKVKRRLLPNNTMLLYALLIOVLVIIESYNV 185  
       ||::|::|::|::|::|::|::|::|::|::|::|:  
Db    266 IsItshwsqvvdvsevlrdgprddeitklrlllIn----- 304  
               |  
QY    186 IPTLKFRKEVAVYYILLFLPFIISLTQLGYTKCKVEINMAACKQAQRNRTITYSOOIESIRK 245
```

Db	305	--LVPRTSMGVL-----LGAISLFC-----	rdkterrykImqgdlvnyad	342
OY	246	DISEPHNDYLIINTSLRIGTENKDLASTEKTYIQOLEKKGHOLDOTRNIGH-LANIQND	304	
Db	343	alrEshEhmKrlhvi-Igllh-IksykqlEduIl-keanypqe--eIgsllgkIksp	395	
OY	305	AVUGISAKTILDAQKKIADVNVVESSKIQLRPEMELDFGTTILSIDNAIAFAFESLNP-	363	
Db	396	vIagflsklnatd--IghlllneseqIpdgsedqvatlltIlglllnaIealIgre	453	
OY	364	---EIQLAFFKKNGSIVETIONSTRKEQOLD-VSKIKFENKSTKGSNRGIGLAKVNHILN	419	
Db	454	pgeIseIvllhIghvIhceyvnddrgIarIdhIdIdkIvstkgsergvglalvkqven	513	
OY	420	YPKTSLOTSNNHNLRFKOLTI	439	
Db	514	I-ggsIaveserpiIftqffIv	532	
RESULT 11				
AAW83369				
ID	AAW83369	standard; Protein; 442 AA.		
XX	XX	AAW83369:		
XX	AC			
XX	DT	17-FEB-1999 (first entry)		
XX	DE	Streptococcus pneumoniae histidine kinase.		
XX	XX			
KW	Streptococcus pneumoniae; histidine kinase; antibacterial; diagnosis;			
KM	Streptococcal; Helicobacter pylori; infection; bacteriostatic; microbial;			
XX	bacteriocidal; ulcer; gastritis; stomach cancer; gene therapy.			
XX	OS	Streptococcus pneumoniae.		
XX	PN	EP881286-A2.		
XX	PD	02-DEC-1998.		
XX	FE	26-MAY-1998.	98EP-0304138.	
XX	XX			

PR 30-MAY-1997: 97US-0048347.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Biswas S, Throup J, Wallis NG, Zalacain M;

DR WPI: 1999-001392/01.

XX N-PSDB: AAV72649.

XX New Streptococcus pneumoniae Histidine kinase polypeptide and
PT polynucleotide - useful as diagnostic reagents and for prevention
PT and treatment of Streptococcal and Helicobacter pylori infections

PS Claim 1; Page 34-35; 44pp: English.

XX The present sequence represents histidine kinase (HK) isolated from
CC Streptococcus pneumoniae. HK polynucleotides and polypeptides are useful
CC for diagnosing susceptibility to diseases by detecting mutations or
CC polymorphisms in the HK gene or analysing for the presence or amount
CC of HK polypeptide expressed in a patient sample. HK PCR probes are
CC useful for diagnosing diseases (especially Streptococcal), and can
CC characterise the stage and the species or strain causing the infection.
CC The HK probes can also determine the response of the infectious organism
CC to drugs. HK polypeptides and polynucleotides are useful for screening
CC for antagonists, agonists and drugs against infectious micro-organisms.
CC HK agonists and antagonists are bacteriostatic and bacteriocidal
CC compounds which can be used in treatment to enhance (agonist) or block
CC (antagonist or antisense sequence) HK activity, therefore treating
CC microbial (especially Streptococcal) diseases, ulcers and gastritis, and
CC stomach cancer caused by Helicobacter pylori. Epitopes of HK polypeptides
CC and polynucleotides are useful immunogens for producing anti-HK
CC antibodies for vaccines to prevent bacterial infections, and HK
CC polynucleotides can be used in genetic immunisation (gene therapy) to
CC prevent infections. HK polypeptides and polynucleotides and their
CC (anti)agonists can prevent adhesion of bacteria to matrix proteins, and
CC are useful for use on wounds and body implants to prevent bacterial
CC infection.

XX Sequence 442 AA;

Query Match 6.7%; Score 145.5; DB 20; Length 442;

Best Local Similarity 21.7%; Pred. No. 0.00019;

Matches 104; Conservative 70; Mismatches 154; Indels 151; Gaps 21;

QY 24 LFSKSNVTLTKKELTFSISNFMIMAVMNVNLFYPAEPLFYALSIYV-NRONSLS 82

DB 6 lftkiflyfs-----lfsvlclhla-----yflfptyshrgctlg 46

QY 83 LNIFFGLPVASDLPFRRAIFFILD-----GTGGIWMGSSITTYMIEFGIALSYLF 136

DB 47 qkataiaagslegkd--rqslegvldysqtsdtkgvkge-----mte----- 87

QY 137 LSVFNVDGRIKDSL-----TKMKVKRRLPMNI--TMELYLLI-----QVLY 178

DB 88 -----dklevkdsldpdtqtslfierevktqdggtmqlfiasmdlqkeagisl 141

QY 179 VIESNVNPTLKFRKFPVYVYLLIFLLISLSQYTKQKQVEMAKQQR----- 231

DB 142 gflpylliaasflisllyavlyartliavpllelkrvtr-----mmdldsqyrlvdskd 195

QY 232 NITQYSGQIESLYKD-----ISFRHDYLNITFSRLGIE 266

DB 196 efgnklqeginslqhlitviadlhekneallqlekmkveflgashelktpasikille 255

QY 267 N-----KD-----LASIKYHQLE-----KTGHQODPTFYNIGHLANIONAV 306

DB 256 nmremigrtykrdqlygvalgydelnhvlgllslssvqelrddretldllqmtgnlvk 315

QY 307 KGIISAKILIAEO-----NKRIAVNEVSSKIQLPMEMLDFTTILSDCNVAIAEAS 360

DB 316 dyallakereidqiansltlqgqylnpsv-----mkl-----llsnlsnaikhsy-- 360

QY 361 LNPETQLAFKKNGSIYFIIONSTRKQIDVSKIFKENYSTKGSNRGIGLVHILEH 419

DB 361 --pgglvtrigerel-fienccseegeklqagtsidnasirvkysgmglfivksilleh 416

RESULT 12

AAV57684

XX AAV57684 standard; Protein: 442 AA.

AC AAV57684;

XX 16-MAR-2000 (first entry)

DE Streptococcus pneumoniae sensor histidine kinase.

XX Antibiotic; growth inhibition; bacterial autolysis; regulation;

XX antibacterial; infection; inflammation.

XX Streptococcus pneumoniae.

XX W09957281-A2.

XX 11-NOV-1999.

XX 06-MAY-1999; 99WO-US09792.

XX 06-MAY-1998; 98US-0073541.

XX (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Novak R, Tuomanen EI;

XX WPI: 2000-062151/05.

XX N-PSDB: AAZ48094.

XX New nucleic acid and peptides, useful as antibiotic peptides

XX Claim 38; Page 128-130; 151pp: English.

XX The present invention describes peptide antibiotics which inhibit the
CC growth of pneumococci, ABC transporter and two-component signal
CC transduction system proteins from Streptococcus pneumoniae. The peptides
CC are useful (especially in the form of a pharmaceutical composition) for
CC the treatment of a bacterial infection or inflammation. Methods from the
CC present invention are also useful for the identification of agents or
CC drugs which are useful in preventing bacterial proliferation or kill
CC bacterial cells e.g. potential antibiotics. The peptides and methods are
CC also useful for providing important epidemiological tools. The peptide
CC are especially useful for the prevention of any disease caused by a
CC bacterium e.g. Streptococcus aureus, Acinetobacter, Enterococcus
CC faecalis, Escherichia coli, Pseudomonas aeruginosa all of which can
CC cause blood poisoning among other ailments. Mycobacterium tuberculosis
CC and Neisseria gonorrhoeae which causes gonorrhoea. The peptides are also
CC useful in the treatment of infections due to Streptococcus pneumoniae
CC bacterial species that causes blood poisoning, middle ear infections,
CC pneumonia and meningitis in humans. The present sequence represents a
CC sensor histidine kinase from the present invention.

XX Sequence 442 AA;

Query Match 6.7%; Score 145.5; DB 21; Length 442;

Best Local Similarity 21.7%; Pred. No. 0.00019;

Matches 104; Conservative 70; Mismatches 154; Indels 151; Gaps 21;

QY 24 LFSKSNVTLTKKELTFSISNFMIMAVMNVNLFYPAEPLFYALSIYV-NRONSLS 82

DB 6 lftkiflyfs-----lfsvlclhla-----yflfptyshrgctlg 46

QY 83 LNIFFGLPVASDLPFRRAIFFILD-----GTGGIWMGSSITTYMIEFGIALSYLF 136

DB 47 qkataiaagslegkd--rqslegvldysqtsdtkgvkge-----mte----- 87

```

QY 137 LSVFNVDIGRLKDSL-----TKMKYKKRLIPMNI--TMLLYVILI-----OVLY 178
      |||||
Db 88 -----dklevdsipldtdrgtstlfieerevktqdgumtqlflasmtdlqkeaqisl 141
      |||||
QY 179 VIESNVIPFLKFRKRVIVVILFLILISPLSQYTKQKQVONEIMAKQAQIR----- 231
      |||||
Db 142 qflpyelllasflslvaylvarclvaplflkrlvtr-----mmldsgvrlrvdskd 195
      |||||
QY 232 NITQVSQOIESLYKD-----IRSPRHQDLNLTSLRGITE 266
      |||||
Db 196 elgnlkeqnslyqnlvleadihhekeaalqlkekveffrgashefkprlaslkille 255
      |||||
QY 267 N-----KD-----LASIEKIVHQILE---KTGHQLODPTRYNIGHLANIONDAY 306
      |||||
Db 256 nmrenigrkddqylgvalgldvdelnhvllslssvgeflrdreldlqmtqlvk 315
      |||||
QY 307 KGLISAKLLEAQ-----NKRIAVNEVSSKIQLPPEMLDPTITLSTICNATEAPES 360
      |||||
Db 316 dyallakereidqdnslthqeylnpsv-----mkl-----lslslsnaikhsv-- 360
      |||||
QY 361 LNEEIQLAEFKKNGSIVFIIONSTKEQIDVSKIFKENVSTKGSNRGIGLAKVNHILEH 419
      |||||
Db 361 --pgyivrigeregel-flenscssegeklagfsdnasrkvkysgmqlfvvksllsh 416

```

RESULT 13

AA069977 standard; Protein; 861 AA.

```

XX AC AA069977;
XX DT 19-NOV-1998 (first entry)
XX DE Staphylococcus aureus histidine kinase.
XX KW Histidine kinase; KdpD; prevention; treatment; diagnosis; vaccine.
XX OS Staphylococcus aureus.
XX PN EP863208-A2.
XX PD 09-SEP-1998.
XX PE 17-FEB-1998; 98BP-0301167.
XX PR 25-FEB-1997; 97US-0039478.
XX PA (SMK ) SMITHKLINE BEECHAM CORP.
XX PA (SMK ) SMITHKLINE BEECHAM PLC.
XX PT Wallis NG.
XX PS Claim 14; Fig 2; 31pp; English.
XX CC The present sequence represents a Staphylococcus aureus histidine kinase.
XX CC The protein is related by amino acid homology to the Escherichia coli
XX CC Kdp histidine kinase protein. The protein can be used to treat an
XX CC individual in need of histidine kinase, while the antagonist is used to
XX CC inhibit histidine kinase. DNA encoding and expressing histidine kinase
XX CC are used in immunological compositions. The active agents are useful
XX CC for preventing, treating, diagnosing and vaccinating against infections
XX CC of the upper and lower respiratory tract, cardiac disorders,
XX CC gastrointestinal disorders, CNS disorders, eye disorders, kidney and
XX CC urinary tract disorders, skin disorders, and bone and joint disorders.

```

CC They can also be used for assaying genetic variation and raising an
 CC immunological response against e.g. Staphylococcus aureus.

XX Sequence 861 AA;

Query Match 6.6%; Score 143; DB 19; Length 861;
 Best local similarity 21.2%; Pred. No. 0.00081;
 Matches 109; Conservative 79; Mismatches 155; Indels 172; Gaps 25;

```

QY 17 TYVFLFLFKSVNTLSKELTLEFST--SNFLIMAVTVNVLNF-----YPAE-- 64
      |||||
Db 414 tlllmflfiig-----llsltwrsffligflaainlvrfnyfftepryfevy 462
      |||||
QY 65 -----PLVFALSIYLNRONSLSLNFIYGLLPVASSDFERRAIFPIDGTQGIYMGSSIT 120
      |||||
Db 463 rfdypitffi-vsllsslltsalqlkqtkqytkkqlyrcdlqlqfnd-----slk 512
      |||||
QY 121 TTYMTIEFAGIALSYLFLSVFNVDIGRLKDSLTMMKRVKKRLIPMNTMLLYLLIOVLYVI 180
      |||||
Db 513 qlyvevnlhnaqyql-----nqllgqslt-----lyvl 541
      |||||
QY 181 ESYNVIPFLKFRKFV-----VIYVLI-----LFLILIS----- 208
      |||||
Db 542 ngskviktkipqnbidntqthegalsvklnerqagatctfpglnkwlipigtspikg 601
      |||||
QY 209 -FLSQYTKQKQVON-----EIMAKQAQIRNITQVSQOIESLYKD-----IRSF 250
      |||||
Db 602 llaigyssqylnpydasllsmelnelavenvlllqtrleslqgerqlthsnflrsl 661
      |||||
QY 251 RHDYLNITLSRLGIE-----NKDLASIEKIVHQILEKTGHQLODPTRY-----NIGHLA 299
      |||||
Db 662 shdirtplttlmgldilvshskdmstlek--eqllv---hsfgesgyllylvtnhlslt 716
      |||||
QY 300 NIQNDKAVKGLSA-----KILEAQNKKAIVNEVSSKIQLPPEMLDPTITLST 348
      |||||
Db 717 kqssnvqklqpylvselveidmlerhkrkrltvsstvnqlfhlidskllqlaln 776
      |||||
QY 349 LCDNAITE--AAFEENPEIOLAEFKKNGSIVFIION-----STKEQIDVSKIFKENVST 401
      |||||
Db 777 llenavkhsdtklnslryasyeq---lefavideqpgalsleeqg-----klfeffyt- 828
      |||||
QY 402 KGSNR-----GIGLAKVNHILEHYPKTSIQ 426
      |||||
Db 829 -gsnkryfkdnqkesmgllqlylvqtl--hkhsnldg 861

```

RESULT 14

AA037397 standard; Protein; 909 AA.

```

XX AC AA037397;
XX DT 14-FEB-2002 (first entry)
XX DE Staphylococcus aureus cellular proliferation protein #1567.
XX KW Antisense; prokaryotic cellular proliferation protein;
XX KW antibiotic; antibacterial; drug design.
XX OS Staphylococcus aureus.
XX PN WO200101955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.

```


Best Local Similarity 19.7%; Pred. No. 0.00027;
Matches 84; Conservative 83; Mismatches 125; Indels 135; Gaps 20;

```
QY 36 KETLTFESINFLIM-ATYMNVNLFYPAEPYFIALSTYLNKNSLSLNIYGLPVAS 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 kdtstaelsiylfiflaqlifnevypkwyllsvivf---sis---yillyvnh 55
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 95 SDFRRATFFIIDGTGIVMGSSITTYMIEFAGIALSYFLSVFNVDIGRLKDSLTKM 154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 nrl--nllmfyl-----llyflicyfvsvhpmislfyysafavpf-tfknyvktc 107
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 155 KVKKRLIPNITMLLYLLIQVLYIESYVNIPTLKRKRVYIYLLIFLLISFLSQYT 214
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 atnlffllmllctillyllynnyfva-----mmvyyvyslilmldnfkmk 153
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 215 KQVQNEIWAQKEAQIRNITQYSGQIESLYKDIRSFHDYINILTSRLGIENKDLASIE 274
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 nreygkel-aeknrhncl----- 171
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 275 KIYHOLEKTGHQLODTRYNIGHLANIONDAVKGILSAKILEAONKTI-----AVNV--- 326
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 -laegerhrigqldht--lghv--faslskseykklidadvkvaellainklsr 225
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 327 -----EVSSTKIOLPE-MELDFTILSIICDNALFEAFESLNPEIQOLAFPKKNGSIVF 378
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 eslnkvrelldvklpsfiseid--sirtvklkadidftfe--nkelagvlsplkgsmly 281
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 IIONSTKEKOIDVSKIFENYSTKGSNNGIGLAKVNHILEHYPKT---SIQTSNHHILF 434
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 ml---Cre-----ainnvikhanaskvhgklktvnnh--- 310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 435 KQLLIRK 441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 klllme 317
```

Search completed: July 30, 2002, 15:14:13
Job time: 1594 sec

OV 12 LITVLTVI--FLLTFSKVSNTISKKELTIESISNFIIMIAVT-----MVNVNLEYPAE 64

```

Db      6  ILLYALVINGLTVIEFRVNGIGLTPRIRKAPLFLGLGIFTFPOFLAVSKVLSYFIE 65
QY      65  PLTFALSTIYLNRQNSLSLNPVGLLPVASSDIFRRALIFFLDGT-QGIY--MGSSIT 121
Db      66  PLUGIGLSLFLNRGLPKKLTILFYGLPMLIVELTFYGVSTFVLPGLGGIYDGDGNEFL 125
QY      122  TYMIEFAGIALSYELFSEVNDIGRLKDSLTKKVAKRLIPANNITMLLYLLIGVLYIE 181
Db      126  LIMIFPCFIVL--VELKMLDDPFLRLREFLDGFOKSLTKITMMAGVLYVMOSLSYLE 183
QY      182  SYNVIFTLFRKFAVIVYLLIFLLISFLSQYKQVQVMEIMAAKEADININITYSQOE 241
Db      184  YEGCIGISTVYRLHLIFLFFPKGICKLPTYLKEKJDELNOEOTLAKROMERSHAIE 243
QY      242  SLVKDIRSRHDYLNILTSRLGIENKDLASIEKLYHQILEKTYGHQLODTRYNTIGLANT 301
Db      244  ELKKEIRSRHDYTNLTSRLGLEEDMDQIKEIYDSVLRDSOKLODNKYDGLRVNI 303
QY      302  QNVAVYGIISAKTIEOKRKTAAVVEYSSKTOLPEMELIDPTIISLIDCAVIAFAEST 361
Db      304  KRDALMSLAKGFITAREKNIVYFVVEPEIOYEGSLDELITVYSLIDCAVIEVSEAS 363
QY      362  NPEIQAFERKNGSIVFTIQNSTEKQIDVSKIFKENYSTKSGNNGIGLAKVNIHLEHY 421
Db      364  QPVAISYAFKLNCAQOETFTIENSIEKEEIDISEIFSGASSKEEGVGLTYVMKIVESH 423
QY      422  KTSLOTSNNHHLFKOLL 440
Db      424  NTMLNTTCOVYFROYLV 442

RESULT 2
US-08-879-941-2
: Sequence 2, Application US/08879941
: Patent No. 6268172
:
GENERAL INFORMATION:
APPLICANT: Mallis, Nicola
TITLE OF INVENTION: NOVEL HISTIDINE KINASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NJ
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/879,941
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10021
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 446 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-879-941-2

```

```

Query Match: 34.6%; Score 752; DB 4; Length 446;
Best Local Similarity 39.2%; Pred. No. 8, 9e-58;
Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;

OY 12 LLYIVLV--ELLEFSKSVNTLSKELTEFSNSFLIMATV-----WVNNVLFYPAE 64
Db 6 LLYLVINGKLVIFEFKVNIGLTFDRIFRAFLTKLELLGIFFTFQELAVSKYLSYFTE 65
OY 65 PLVFALSTLYMKNRNSLSINIFYGLPVASSDLFERRAIIFFILDTG-QGIV--MGSSIT 121
Db 66 PLFGIGSFLLTGLPKRIILIFYGLFPMILVELFYRGVSYPVFLFOCGIVDDGDIPL 125
OY 122 TYMEFGALISTLPSFVFNVDIGLKNLSLTKMKYKKRLPMNTMLLYLLQVLYVE 181
Db 126 LMFVCEVLV-VFLKMLDYFTLRKREPLDTGFKSLTKIKNMAMATYLVWQSYLSE 183
OY 182 SYNVVPLFKRKRVIVIVYLIIFLLISLSQYTKQKVOVNEIMACKEQIRNTQYSQTE 241
Db 184 YEGGISTVYRHLILVLYLLFPMGSIKKLDYLLKELQEELENOQTLYRDMRYSRHE 243
OY 242 SLXKDISEFHDYILNITLSLRLGINKKLASIEKIKYHOILEKTGHOLQDRYVNIIGLANI 301
Db 244 EIKKEHSFPHDYTNLTJSLRGLDEEDMEQIKELYSVLRDSQKQKQDKYDLGLRVNI 303
OY 302 QNNAVGCILSARKILEQNKRIAVNVEYSKIQLEPMELOPITLITSLCNALAEAFESL 361
Db 304 RDRRLSLSLGKRIKREKNIVNVEPELVQEBMSLDTLTVLISLCNALAEVSAEMS 363
OY 362 NPETQIAFFKRNKSIVFIITNSFKRKQIDVSKIRFKNYSYTKSGNSRGIAGLAKVNHILEHP 421
Db 364 QPHYSIAFLKNGAQEFTFIENSIKEGIDISEIFSFGASSGKEERGGLTYVAKIYESHP 423
OY 422 KTSLOTSNHHLEFKOLL 440
Db 424 NTNNTTCQNVQVRYLVTV 442

RESULT 3
US-09-747-116-2
Sequence 2, Application US/09747116
Patent No. 6348340
GENERAL INFORMATION:
APPLICANT: WALLIS, NICOLA
TITLE OF INVENTION: NOVEL HISTIDINE KINASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 997 Lenox Drive, Building 3, Suite 210
CITY: Lawrenceville
STATE: NO
COUNTRY: USA
ZIP: 08543
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASTSEO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/747, 116
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/879, 941
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GMI0021
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259

```


TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 446 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-09-747-116-2

Query Match 34.6%; Score 752; DB 4; Length 446;
Best Local Similarity 39.2%; Pred. No. 8,9e-58;
Matches 172; Conservative 86; Mismatches 169; Indels 12; Gaps 5;

QY 12 LLYPLVYL--FLFLFSKSNVTLSSKKELELSISNPLIMAVT-----MNVNLFYPAE 64
DB 6 ILXALVINKLIVIPFKNGIGLTFDRIFKFLKGLIGLITFTFQPLANSKILSTFIE 65
QY 65 PLVFALISYLNQNSLSINIFGLPVASSDLFRRATIFLLDGT--OGIV--MGSSITT 121
DB 66 PLFGIGLSFLLRGLPKLILFYGLEPMLIVLEFYRGVSFVLPFGGIVDGDGNPIFL 125
QY 122 TYMIEFAGIALSTFLSVENVOIGRLKDSLTKMKVKRLIPNITMLLYLLIOVLVYE 181
DB 126 LIMFVCEYVL--VFLKMLDYDTRLRREFLDTGFKSLTKINMAGAYILVMOSTYLE 183
QY 182 SYNVIPPLKPKKRVIVYVLLFLILISLSQYTKOKVONEIMAOEAOIRNITQYSGOIE 241
DB 184 YEQIGISTYRHLILVFLYLLFFMGIGIKLDYFLKELQELNOEOTLRHYRMERSRIIE 243
QY 242 SLKYDISFPHDYLNTISLRGLENKDLASIEKTYHQLEKTHGLODTRNIGHLANI 301
DB 244 ELKYEISFPHDYNLTLSRLGIEEDMDQKEITDYLRSOSKIDNDKIDGLRVNI 303
QY 302 QNDVAVGILSAKLEAKNKIAVNEVSSKIDLPKEMELDITLISLIDNAIEAEPESL 361
DB 304 RDRALKSLAGKRIKAREKNIVNVEPEIOWEGSKLDFLIVSLICDNLAEVSAEAS 363
QY 362 NPEIOLAFPKNGSIYFIIONSTKREKIDVSKIFKENTSTKSGNIGIAGAKVNHLEHY 421
DB 364 OPHYSIAFLKNGAOEFILNISKEGIDILISLFSFGASKSKEBVGGLYIVMKIVESH 423
QY 422 KTSLSQTSNHHHLFKOLLIT 440
DB 424 NFNINNTTCQNOVFROYLV 442

RESULT 4

US-08-858-207A-327
Sequence 327, Application US/08858207A
Patent No. 6348328

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328e1 Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A

FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 327:
SEQUENCE CHARACTERISTICS:
LENGTH: 108 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-327

Query Match 11.9%; Score 258; DB 4; Length 108;
Best Local Similarity 48.1%; Pred. No. 1.6e-15;
Matches 50; Conservative 22; Mismatches 32; Indels 0; Gaps 0;

QY 337 MELDFITLSTICDNATAEAFESINPEIOLAFPKNGSIYFIIONSTKREKIDVSKIRK 396
DB 1 MSLDPLFVSLICDNALAKASASQPHVSIAFLKNGAOEFILNISKREGIDISEIRS 60
QY 397 EWSYTKSGNRGTGLAKVNHLEHYPKTSLSQNSHHHLFKOLLIT 440
DB 61 FGASRSGERGVLTYMKIVESHENITNITTCQNOVFROYLV 104

RESULT 5

US-08-858-207A-326
Sequence 326, Application US/08858207A
Patent No. 6348328

GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Nicholas, Richard
APPLICANT: Stodola, Robert
TITLE OF INVENTION: No. 6348328e1 Compounds
NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478

TELEX: 210 001 0222
INFORMATION FOR SEC ID NO. 2.

Query Match 6.78; Score 145.5; DB 4; Length 442;

```

: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 618

```

TYPE: PRF
ORGANISM: Staphylococcus warneri
US-09-299-378-4

Query Match 5.4%; Score 117; DB 4; Length 618;
Best Local Similarity 21.4%; Pred. No. 0.038;
Matches 116; Conservative 76; Mismatches 187; Indels 164; Gaps 27;

```

QY 8 LSNGLTLYTVL-----FLTLFESKSVNVTLSKKELEFLFSISNLI-----MIAV 52
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 92 LNTLTITKTKVLPACGCGISIAFIEIENKQISKNSIDF-INNVAVIDVSNEMINE 150
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 53 TAWN-----VNLTPAEP-LYFI-----ALSTIYNRONSLSN-----84
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 151 TKINILTMYYLYNDISIFDKVKPNYIAIDFVQEKKEFENLFENFSQONFINNFEN 210
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 85 --IFGGLPVASSDLF-RRAT-----IFELIDGTG 112
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 211 FDIYIGNPPYVS--LYGRRAINKSEDKRQFLIRNYDFIPKNNKNGKNTYMFIEINGLOL 268
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 113 IVMGSSI-----ITTYMIEFAGIALSYLFLSVFN-----VDIGRLKDSL 151
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 269 LKKNGLTFEIVDTLLESSESIRKYLETAIIKOLDINIKAFSDVVSQIIISLKNAS 328
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 152 TKMKV-----KKRLIMNITMLL---YLLI---QVLVIES--YNAVPTLK--PRKF 194
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 329 NKEAIVSTKDMONNTTQINDIMLHDKFYEPNSTDKKINSILEKYNNKSDDELQYTFPPK 388
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 195 VVIYVLLIFLISFLSOYTRQKQVONEIMAKOEAOIRNTOYSOIES-----LYK 245
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 389 ELRTSTMLNMESSEFVKDY-KPEIDFHVMPYKCG-AKNLSEPFQNMHSNHYFYDALOK 446
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 246 DIRSRHRYL---NLTLRLIGIENKDLASIEKIY-----HOLEKQHGLQDTRNIGH 297
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 447 KINSDHEELKKGKIKKKRIGLGNLEFKPKLFIQSANAKLATDGCNMSNNSLYI 506
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 298 LANIONDAVKGILSAKILEAQNKKIAVNEVSSKIQLPEMELLDTITLISILCDNAIEAA 357
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 507 LSKATND-IKDIIMLKITCAOLNS-----ELLTFIL-----TNRIRK 544
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 358 FESLMEIOLAFKKNGSIVFIIONSTKEKQIDVSKIFKENYSGKSNRIGIAGAKVHIL 417
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 545 AAGKQOLSLDK--TIPLCFNEINSKLL-----IFAENATKKNNELSSLEKINQII 557
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 418 EHY 420
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 598 YKY 600

```

RESULT 10
US-08-963-901-2
Sequence 2, Application US/08963901
GENERAL INFORMATION:

APPLICANT: Wallis, Nicholas G.
APPLICANT: Traini, Christopher M.
APPLICANT: Kosmatka, Anna L.
APPLICANT: Shilling, Lisa K.
APPLICANT: Warren, Richard L.
TITLE OF INVENTION: NOVEL HISTIDINE KINASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,901
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Dickinson, Todd O
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10058
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 457 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-963-901-2

Query Match 5.2%; Score 114; DB 4; Length 457;
Best Local Similarity 19.6%; Pred. No. 0.045;
Matches 96; Conservative 76; Mismatches 167; Indels 152; Gaps 24;

```

QY 38 LTFESISNELLIMAVTNVNLLEY-----PAEPLFYALSI 73
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 9 IAVISITVELSALISFLVNNVYHYNLKASNDAKIMTKLEARQEOSAKPTHIQOYK 68
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 74 YLNONSLSNI-----FYGLPVASSDLFPRAT-----F 104
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 69 HLGQNVYIMVVDKGRKTFG--EPREDLSQALNNVNNKDYGINDKPFALEVGF 127
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 105 F--IIDGTQIVMGSSIITTYMIEFAGIALSYLFLSVENVYDGR-LKDSLTPMKYKKRLI 161
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 128 FDNVDNTVGI-----NEKTKDGSIAV-----FMRDIOETFSERTFLAV-----168
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 162 PMNTMLLYLLIQLVLYIESVNIPTLKRFKRVYIYLLIFLISFLSOYTRQKQVONE 221
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 169 ---LMLLFLSIS-LVIASYSIIRPVKKL-----LATELLIDGDETPRKQTRKE 218
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 222 I-----MAKQEAQIRNTOYSQ-----TESLYKDY-----RSFRHRYL 255
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 219 IGTLYQHFNKMRRESLQGVQDQKHQFVONVSHETKPTLTHIHLLSELQOTSOKTLRQYI 278
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 256 N---ILTLRLIGIENKDLASIEKIYHOLEKQHGLQDPRVNYNGHLANIONDAVKGILSA 312
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 279 NDIYITITQLSGLTTELLLSLSDNHO-----HLFPDKIOVDL-----IKDIINH 325
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 313 KILEAQNKKIAVNEVSSKIQLPEMELLDT-----TLISILCDNAIEAAEFESLNPEI 365
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 326 EQFADEKSLIT-----LADLESINFLGNQRLHQAISMLINAIK--YTDVGAI 374
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 366 OLAFKKNGSIVFIION--STKEKQIDVSKIFKENY--STKGSNDRIGIAGAKVHILE-HY 420
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 375 DALHSHNNIITFISNDGSPISPOAE-ARLEFEKYVSHDNSNGSLATIKTSITELHH 433
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 421 PRTSLQTSNHH 431
   : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 434 GIIQFTQSNEX 444

```

RESULT 11
US-08-568-985-2
Sequence 2, Application US/08588985
Patent No. 5777094
GENERAL INFORMATION:
APPLICANT: Michiyuki MATSUDA et al.
TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN

NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,985
FILING DATE: January 19, 1996
CLASSIFICATION: 336
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1865 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: spleen cell of homo sapiens
US-08-588-985-2

Query Match 5.2%; Score 113; DB 1; Length 1865;
Best Local Similarity 20.1%; Pred. No. 0.39;
Matches 95; Conservative 90; Mismatches 208; Indels 80; Gaps 21;
QY 21 LLEFESKSVNVTLSKKELEFISNFIIMIAVTMNVNLFYPAEPIYFALSIYLN-RQ 78
DB 772 LLOLFRSINDMMSMSDQIV-RVKGAAKLYPLTIYNDVKLYVDEPKELSKMTEFILLNVP 830
QY 79 NSLSLNIIFGILPVASSDLFRRALIFILIDGTGIVMGSSITTYM-----IFRAGIALS 133
DB 831 GLITIOKLYCLIEIYHSDLFTHDCREIL-----LPMWTDQKLYHLEROEDLEACCOLLS 885
QY 134 YFLFSVFNVDIGRLKSLTKMKVKKRLIPMNTML-----LYLLIOVLVYIES 182
DB 886 HILEVLYRKVDGPTGRHV-QIIMEKLRVTNRTVISMGDSLEIGNFVACMTAILRQMED 944
QY 183 YNNIPLK-FRFVVIYVLLIFLISLSQTKO-----KVQNEIMAQKEAQIRNITO 235
DB 945 YRAHALKTKTGKMTGVNDFLMEFTIMERNLIGKNVYPPDWVIMN-VWQNKVFLRAINO 1002
QY 236 YSQOIESLYADIRSF---HDYINI-----LNSLRIGIENKDLASIERI---YHOILEKT 284
DB 1003 YADMLNKFLDQANFELQIMNNYFHLAVAFLOESLOJENFSSAKRAKILNKGDMRROI 1062
QY 285 GHLODTRVYIG-HLANIONDAVKGLSKIL-EAONKKAIVN-----EVSXKIQ 333
DB 1063 GFELRDMVYVGOHKIKRIPENAGPLLEWTLIPETELKRAPIPFDMAGCEFHSTNSQ 1122
QY 334 LPEMELDFTTILSLCDNAIEAA-----FESLNPEIQ-----AFKNGSIVFIION 382
DB 1123 MFEREI---ITKL---DHEVEGGRGDEQYKYLFDKILHECHRKHKYLAKEGFEVVKLV 1175

QY 383 STREKQDVSKIFKENVSTKSGNSRGICLAVNHILEHYPTSLQTSNHHHLER 435
DB 1176 RIMERILDYRTIMHD-----ENKENHMSCTVNLNVEYKEIEREMVYIRYLVK 1222
RESULT 12
US-08-971-988-2
Sequence 2, Application US/08971988
Patent No. 5786461
GENERAL INFORMATION:
APPLICANT: MICHIOKI MATSUDA et al.
TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,988
FILING DATE: 17-NOV-1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/588,985
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1865 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: spleen cell of homo sapiens
US-08-971-988-2
Query Match 5.2%; Score 113; DB 1; Length 1865;
Best Local Similarity 20.1%; Pred. No. 0.39;
Matches 95; Conservative 90; Mismatches 208; Indels 80; Gaps 21;
QY 21 LLEFESKSVNVTLSKKELEFISNFIIMIAVTMNVNLFYPAEPIYFALSIYLN-RQ 78
DB 772 LLOLFRSINDMMSMSDQIV-RVKGAAKLYPLTIYNDVKLYVDEPKELSKMTEFILLNVP 830
QY 79 NSLSLNIIFGILPVASSDLFRRALIFILIDGTGIVMGSSITTYM-----IFRAGIALS 133
DB 831 GLITIOKLYCLIEIYHSDLFTHDCREIL-----LPMWTDQKLYHLEROEDLEACCOLLS 885
QY 134 YFLFSVFNVDIGRLKSLTKMKVKKRLIPMNTML-----LYLLIOVLVYIES 182
DB 886 HILEVLYRKVDGPTGRHV-QIIMEKLRVTNRTVISMGDSLEIGNFVACMTAILRQMED 944
QY 183 YNNIPLK-FRFVVIYVLLIFLISLSQTKO-----KVQNEIMAQKEAQIRNITO 235


```

?      TOPOLOGY:  linear
us-09-081-686-2

Query Match          5.0%; Score 108; DB 4; Length 446;
Best Local Similarity 18.3%; Pred. No. 0.15; Mismatches 133; Indels 218; Gaps 23
Matches 96; Conservative 78;

QY      36 KETLIFSISNFIIMAVTNVNVNLFYPAPRPLTFIALSTYLNRONSLSINIFYGLHPASS 95
      | | : : : : | : | : | : | : | : | : | : | : | : | : | : |
Db      6 KLTNPLRSFALIGVGVLVHAIYLFPPFYIDLE--GKRNESARVETELTKTKTS 62

QY      96 DLRRRAIFILDTGTGIWGSIIITTYTIERAGIALSY---LFLS----- 138
      | | : : : : | : | : | : | : | : | : | : | : | : | : | : |
Db      63 D-----ELPSLQSYSKSLTISALNRDIVDKRLPL 93

QY      139 VFNVNDI--GRLKDSLTKMKVKRRLIPMNI--TMLLYLLIQVLVYIESYNYLPTLKRKYV 195
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      94 VHDLDIDKGLSLYIV-----MLDMSVSTADKQVGVQVGVHGVYV-----KEAKNI 140

QY      196 VIYVL-ILFLLISF---LSQYTKQVQNET-----MAQKEAQIR-----NIYQ 235
      : : | | | : | : | : : : : | : | : | : | : | : | : | : | : |
Db      141 LLVLPYTFVLVTAISFVFSYFETKRLNPLFIYSEVTSKMQDDLDNIRFEDSKRDEVGE 200

QY      236 YSQQIESLYKD-----IRSFHDYLLNLTSLRLGIEN--- 267
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      201 VGRQINCMYEHLLKVTHELESRNEQIVKQNGKVSFVAGASHELKTPLASRLILEMNH 260

QY      268 -----KDLASIEKTYH--QILEKTHQ-----LQD 290
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      261 NIGDYKDPKYIAKSIINKIDQSHLLEEVLESSEKFOEWTECREPVLGVKPYLVLDLSRYOE 320

QY      291 TRYNIIG-ILANIQNDVAVKGLSARILDEAQNKIIVAVVEYSSKIQLPEMELDFTYILSL 349
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      321 IAHSIGVTIENQLTDAVTVVMSLRALD-----KVLTLN 353

QY      350 CDNAIEAAFSNLPEQLAFEFKNGSIYITON---STKE-----KQIDVSKIRKENV- 399
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      354 ISNAIK-----YSDKNGRVLIISQDGYLSIKNTCAPLSDQLEHLRFDIFYHS 400

QY      400 ---STKGSNRGIGLAKVNHILEHP-----KTSLOT 427
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      401 QIVTDKDESSGGLGIYVSNLSEYQMDRYRLPYEHGMEFKISLQF 445

Search completed: July 30, 2002, 15:14:44
Job time: 520 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 30, 2002, 15:15:18 ; Search time 29.21 Seconds
(without alignments)
1450.715 Million cell updates/sec

Title: US-09-833-017-4

Perfect score: 2173

Sequence: 1 MNEALMILSNGLTYLVLE.....KTSIQTSNHHLPKOLLIIK 441

Scoring table:

BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR_71:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	755	34.7	446	2	H97929 histidine kinase (
2	749	34.5	446	2	D95061 sensor histidine k
3	455	20.9	447	2	F96909 histidine kinase-1
4	427.5	19.7	441	2	B95261 probable sensor hl
5	423.5	19.5	441	2	G98126 histidine protein
6	418.5	19.3	431	2	AC1081 sensor histidine k
7	416	19.1	431	2	AD1438 sensor histidine k
8	379	17.4	423	2	S56480 agcC protein - Sta
9	372	17.1	423	2	D89995 accessory gene reg
10	362	16.7	428	2	G56273 protein-histidine
11	348	16.0	448	2	S57909 probable histidine
12	223.5	10.3	452	2	B97095 membrane associate
13	184.5	8.5	433	2	AB1513 two-component sens
14	174	8.0	542	1	E66600 two-component sens
15	171.5	7.9	532	2	E83699 two-component sens
16	163.5	7.5	657	2	D81308 Probable Integral
17	155.5	7.2	533	2	C55521 vira protein - Clo
18	154	7.1	540	2	A70009 two-component sens
19	151.5	7.0	921	1	G71705 alkaline phosphata
20	151	6.9	543	1	D65222 hypothetical 60.6
21	151	6.9	543	2	C91267 two-component sens
22	151	6.9	543	2	H86107 probable 2-compone
23	150.5	6.9	543	2	AH1023 two-component sens
24	148.5	6.8	473	2	G97085 sensory transducti
25	146	6.7	1447	2	F82909 hypothetical prote
26	145.5	6.7	442	2	C95070 sensor histidine k
27	145.5	6.7	442	2	A97938 vncs, histidine k
28	143	6.6	538	2	G84129 two-component sens
29	143	6.6	885	2	D90000 sensor protein kdp

30	142	6.5	824	2	A86783 glycosyl transfera
31	138.5	6.4	363	2	B89907 hypothetical prote
32	137	6.3	428	2	S32935 kinb protein - Bac
33	136.5	6.3	1008	2	F71727 acriflavin resista
34	136	6.3	733	2	C89940 hypothetical prote
35	136	6.3	754	2	G70124 hypothetical prote
36	135.5	6.2	542	2	E90604 hypothetical prote
37	135.5	6.2	569	2	B70113 hypothetical prote
38	134.5	6.2	553	2	AD0579 sensor kinase dph
39	134	6.2	565	2	G64532 methyl-accepting c
40	133.5	6.1	538	2	D82180 methyl-accepting c
41	131	6.0	564	2	F71975 methyl-accepting c
42	130.5	6.0	714	2	C90100 hypothetical prote
43	130	6.0	590	2	G90127 hypothetical prote
44	129.5	6.0	339	2	C96927 sensory transducti
45	129.5	6.0	532	2	H83993 two-component sens

ALIGNMENTS

RESULT 1
H97929 histidine kinase (EC 2.7.3.-) [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: H97929
R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
K:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A978/2; M01D:21429245; PMID:11544234
A:Accession: H97929
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK9268.1; PID:gl5458033; GSPDB:GN00174
C:Genetics:
C:Keywords: phosphotransferase

Query Match	34.7%	Score 755;	DB 2;	Length 446;
Best Local Similarity	39.1%	Pred. No. 9.5e-39;		
Matches 175;	Conservative 88;	Mismatches 171;	Indels 14;	Gaps 5;
QY	1	MNEALMILSNGLTYLVLEFLFESKVSNTSKKELTFSINPLIMAVT-----MV	55	
DB	1	MAINMILKALVINGLETIV---TFPKVDGIGTFDRIFKAFILKFLGIIITTFPOFLAV	56	
QY	56	NYNLYPAEPLFYALSLYLRNONSLSINIFYGLLPASSDERRAIFIELDGT-OGIV	114	
DB	57	SKYLSFLEPTEIGLSIFLLRGLPKKILIFPGJPMILVELRGVSYPVLPFLGQIV	116	
QY	115	-MGSIIITTVIERFAGALSYFLSPVNDIGLTKSLKMKVKKRLIPMNTMLLYL	172	
DB	117	DGDGAPFILLMIEFCFVL--VELKMDIDPFTLRREPLDTGFQSLTIINMAAYIL	174	
QY	173	LIOVLVIESYVNPFTLFRFVVIVYLIFLLISPLSOYTRKQVONEIMAKREAQIN	232	
DB	175	VMSQSYLEVEGQISTVVRHLIVFYLLFPMGSIKKLDLYLEKLOEELNOCOTLRVD	234	
QY	233	TRYOSQIESLYKDIRSRHDYLNITSLRGINKDLASTIKRYQILEKTGHODPR	292	
DB	235	KRRYSKHIELEYKTRSRHDYTNLISLRGIEEDMEQIKELYSVLRDSOKIQDNK	294	
QY	293	YVIGHLANTQNDVAKGISANILEQRONKKAIVNVESSKIQLPBMLIDFTITLSLCN	352	
DB	295	YDLGRIVNIRDLAKSLAGFKFAKREKNIYVNEVPEEIQVBSMLDPLTIVSILCN	354	
QY	353	ATFAAFESLNPEIQAFPKKNSIYFIITQNSTKQIDVSKIFRENVSTGSRNGIGLAK	412	

```
Db 355 ALEASAEQPHVIAFLKGAQETITENSKEEGIDISLIFSGCASKGEEKRGVGLT 414
OY 413 VNHLEHYPKTSLOTNSHHHLFKOLLIT 440
Db 415 VMKIVESHPTNTTCQNOQVROVLTV 442

RESULT 2
D95061
Sensor histidine kinase BHPH, probable [imported] - Streptococcus pneumoniae (strain TIGR)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: D95061
R:Jettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: D95061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-446 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74685.1; PID:g14972002; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0527

Query Match 34.5%: Score 749; DB 2; Length 446;
Best Local Similarity 37.8%; Pred. No. 2, 2e-38;
Matches 170; Conservative 89; Mismatches 173; Indels 18; Gaps 4;

OY 1 MNEALMIL-----SNGLLTYLTVLFLEFSKVSNTLS-----KKEFLFSISNFMIA 51
Db 1 NMIAWILLTYLTNGL-----EIVIFKVDGIDLTFRKIFKALITLAFVYVMS 52
OY 52 VMVWVNVNFEPAEPLFYFALSIVLRONSLSLNFYGLLPVASSDFRRAITFEFLDGT- 110
Db 53 YIVGVNVLYSTEMEPLXIGLSFLLRGLPKKLPFLFYGLPMILVNFYGVGYFVPLFG 112
OY 111 OGIVNGSSITTYWIERAGIALSTFLSVNNDIGRLKDSLTKMKYKRLIMNTMLY 170
Db 113 OGQVVDGYSTFGCLIIIFNFELSLAFKWLVDYDFTSLRKEILDKAPQKSLTGINMIGY 172
OY 171 YLLIOVLVIESYNYIPTLKRFKRVIVYLFLILISFLSYTKOKVONEMIAQKQAI 230
Db 173 YIVMSLSFFFEKESIOSKTVRHILVFLFFMGVYKKLDTYLRKEXLERLEQDALRY 232
OY 231 RNITQYSCQIESLYKDIRSFHDYLTLSLRGIEKNDIASIEKTYHQILEKTHQLOD 290
Db 233 ROMEYVSRHIEELKVEVSRHDYLTLSLRGIEEDDEQIKVEYGVGLDKSSOKLON 292
OY 291 TTYNIGHLANIONDAVKILISAKILEAQNKKIAVNEVSSKIQDPMELDLFTILSLIC 350
Db 293 NKYDGLRVNIRDAKAKSLAKGLKARKDNIVFNVEPEIQVEGMSLLDLFTIVSLIC 352
OY 351 DNALTAPEBSLPEIOLAFKKNISYVITTONSTKEQIDVSKIFKENTSTGSGNRGIGL 410
Db 353 DNALTAPEASVQPHVIAFLKGAQETITENSKEEGIDISLIFSGCASKGEEKRGVGL 412
OY 411 AKVNHLEHYPKTSLOTNSHHHLFKOLLIT 440
Db 413 YVMKIVESHPTNTTCQNOQVROVLTV 442

RESULT 3
P96909
histidine kinase-like Arpase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #extl_change 14-Sep-2001

C:Accession: P96909
R:Nolling, J.; Brenner, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: P96909
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-447 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78065.1; PID:g15022902; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0080

Query Match 20.9%: Score 455; DB 2; Length 447;
Best Local Similarity 27.4%; Pred. No. 1, 6e-20;
Matches 124; Conservative 96; Mismatches 163; Indels 70; Gaps 10;

OY 11 GLTYLTVLFLEFSKVSNTLSKKEFLFSISNFMIAVMVNVNFPYAEPLYPFA 70
Db 33 GMIVTTEIISLIFLM-----LGMQGYTVA-----YILI 61
OY 71 LSTVLRONS---LSLNFYGLPVAAS-----DLFRRAITFEFLDGTGIVMGSSIT- 120
Db 62 PCIEFWLQTSVYLSIS-----LTVAGCLIIYWDYFLSVLYVFLVQSVYKMDVSIH 116
OY 121 -TTYWIERAGIALSTFLSVNNDIGRLKDSLTKMKYKRLIMNTMLYLLIOVLYV 179
Db 117 WTFPCIEFYGIVYSRTIKKVRD-----KFKAYNGHLGTLGVVAAGTLIMLFEV 167
OY 180 ISYVNI-----PDLKRFKRVIV-YLILFLLISFLSYTKOKVONEMIAQK 228
Db 168 FNTNVNVPAPNNSVYNSMTMRKIGVLFESYAILLIIVITILRGLIKEME--LKSKE 224
OY 229 QIRNITQYSCQIESLYKDIRSFHDYLTLSLRGIEKNDIASIEKTYHQILEKTHQ 288
Db 225 EFQSLQDYTNKLEKLRKMRGFRHDYLTLSLRGIEKNDIASIEKTYHQILEKTHQ 284
OY 289 OOTRWIGHLANIONDAVKILISAKILEAQNKKIAVNEVSSKIQDPMELDLFTILSI 348
Db 285 KSNMFKIGLQNIPELPEIKGMFSKAIIRAOETGIDYIDVAESIKSFNNEIILSVI 344
OY 349 LCDNALTAPEBSLPEIOLAFKKNISYVITTONSTKEQIDVSKIFKENTSTGSGNR 408
Db 345 LIDNALTAPESEKCDRSMKVAIVNDRKSVYIIVNNVEELIPYIKLYKGFSTKGNRGI 404
OY 409 GLAKVNHLEHYPKTSLOTNSHHHLFKOLLIT 441
Db 405 GLSNLKDIITGKRPVMDLTVIDEQFOIIDIK 437

RESULT 4
B95261
probable sensor histidine kinase Comp [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: B95261
R:Jettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95261
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK76283.1; PID:g14973747; GSPDB:GN00164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
```

A:Gene: SP2326

Query Match 19.7%; Score 427.5; DB 2; Length 441;
 Best Local Similarity 25.9%; Pred. No. 7.6e-19;
 Matches 119; Conservative 99; Mismatches 188; Indels 53; Gaps 12;

QY 12 LLYLVV--FLLEFS--KVSNTLSKELTLESINFLMIAVM-----YAVN 58
 DB 3 LFGVTVVHFLIISHSYRLICKGRINRELKELVFGATVLLLEIYEFSEYLLYDKIGIE 62
 QY 59 LFYAPPLFYFIALSTYL-----NRONSLSLNI FYGLLPVASSDLFRRAIFFLDGTG 112
 DB 63 RF-----LPLGLSYFRMMKQYERDRGLFLLSLLESTHNL--SVFSSITGDNF 115
 QY 113 IV---MGSITTYMIEFGIALSTYL---LSFVNDIGRLKSLTKKKVKRLIPANIT 166
 DB 116 VLQYHFFEFVTVVLYEVLKILYFHELYAFPKDY-----LYPLKK 160
 QY 167 MLYVLLIOVLVIESYVNIPLK--FRKRVIVYLLFLILISLQYTKQVONE--I 222
 DB 161 VFALLLLHIVSFVS--DMVSTIKHLNSFGSILSIVFSLILFFPANKSHQOMEKEIA 218
 QY 223 MACKQOIRITQYSOQIESLYKDIRSEPHDNLITSLRGIEKKDLASEKTYHQLE 282
 DB 219 LKQKFEQKHLQNTDEIVGLTNEIRGFHDYAGMLVSGMADSGNLOEDRIYNEVLV 278
 QY 283 KTGHOLODRYRNIGHLANTQNDAYKIGLSAKILLEAQNKKIYANVSVSKLOPPEMLDF 342
 DB 279 KAHKKHLSDKYTFYFDLNTNEDSKRELVAOSIVYARNNGVEFTELEVKTIKPLELIDL 338
 QY 343 ITLITSLCDRAIFAFESINPEIOLAFFKNGSIVETIIONSTREKQIDVSKIFPENYSTK 402
 DB 339 VRIMSVLLNNVAGSADSTKQKQEVAVIKMETEYIVIONSKMTMTPSGDLFALGFSTK 398
 QY 403 GSNKIGLAKVNHILEHYKTSLSQSNHHHLFKOLLITK 441
 DB 399 GRNRGVGLNNVKELDKYNNIILEFMEGSTFROIIRK 437

RESULT

5 histidine protein kinase (EC 2.7.3.-) [imported] - Streptococcus pneumoniae (strain R6)
 G98126
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: G98126

R:Authors: J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; Ee, R.; Leblanc, D.J.; Lee, L.N.; Leikowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M.Y., P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.; A:Title: Genome of the bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: G98126
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-441 <CUR>
 A:Cross-references: GB:AE007317; PIDN:AAL00844.1; PID:915459750; GSPDB:GN00174
 C:Genetics:
 A:Gene: cmd
 C:Keywords: phosphotransferase

Query Match 19.5%; Score 423.5; DB 2; Length 441;
 Best Local Similarity 25.9%; Pred. No. 1.3e-18;
 Matches 116; Conservative 84; Mismatches 171; Indels 77; Gaps 10;

QY 33 LSKRELLESINFLMIAVMNVLFRAPPLFYFIALSTYLNNONSLSNIFGLLPV 92
 DB 28 INKRELVEGAYVLLLEIYFDF-----PLVI---LVLD-----GATIERFPL 68
 QY 93 A-----SSDLFRRAIIFILDTGQIVWGSIIITTYMIEFGIALSTYFLSV 139

DB 69 GLYSFRMMKQYERDRGLFLLSLLESTHNL-----VTSSTGDNFVQIHFFP 124
 QY 140 FNVDIGRLKSLTKKKVKRLIPNNITMLLYLLIOVLVIESY----- 183
 DB 125 FVVTV-----LTFVVLK-----IYFHELYAFDEDEIVPLKKVFFALLLH 169
 QY 184 -----NVIPLK--FRKRVIVYLLFLILISLPSQYTKQVONE---IMACKQOIRNI 233
 DB 170 IYSPVSDVSTYKHLNSGSLSSIVFSLILFFPANKSHQOMEKEIAKQKFEQKHL 229
 QY 234 TQYSOQIESLYKDIRSEPHDNLITSLRGIEKKDLASEKTYHQLEKTHQLODXY 293
 DB 230 QNYDEIVGLTNEIRGFHDYAGMLVSGMADSGNLOEDRIYNEVLVANKHLSDXY 289
 QY 294 NIGHLANTQNDAYKIGLSAKILLEAQNKKIYANVSVSKLOPPEMLDFTTITICDNA 353
 DB 290 TYFDLNTNEDSKRELVAOSIVYARNNGVEFTELEVKTIKPLELIDRLVMSVLNNA 349
 QY 354 IEAFESINPEIOLAFFKNGSIVETIIONSTREKQIDVSKIFPENYSTKSNRGILAKV 413
 DB 350 VEGSADSYKKQEVAVIKMETEYIVIONSKMTMTPSGDLFALGFSTGRNRGVGLNV 409
 QY 414 NHILEHYKTSLSQSNHHHLFKOLLITK 441
 DB 410 KELDKYNNIILEFMEGSTFROIIRK 437

RESULT

6 sensor histidine kinase (AgrC from Staphylococcus) homolog lmo0050 [imported] - Listeria sensor histidine kinase (AgrC from Staphylococcus) homolog lmo0050 [imported] - Listeria
 AC1081
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC1081

R:Authors: Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussuguet, O.; Entlian, K.D.; Fsihi, D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kreft, J.; Kunz, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ok, C.; Schlueter, T.; Simoes, N.; Tjelle, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AC1081
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-431 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC98265.1; PID:916409409; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0050

Query Match 19.3%; Score 418.5; DB 2; Length 431;
 Best Local Similarity 26.8%; Pred. No. 2.6e-18;
 Matches 125; Conservative 88; Mismatches 142; Indels 111; Gaps 15;

QY 21 LFLFSKVSNTLSKELT--LFSINFLMIAVMNVLFRAP--PLYFIAL-----S 72
 DB 4 LMAITIGITGIFALQILLTNKVSFKESGLVITAMALAPFLIYQWMSMIFLYIFVSA 63
 QY 73 IYLNRM-----SLSNIFGLLPVASSDLFRRAIIFILDTGQIVWGSIIITTYMIEF 127
 DB 64 LYMKRNKVVVSAITLVII--LITISDS-----IVGFIL-----VPG----- 99
 QY 128 AGIALSYFLSFVNDIGRLKSLTKKKVKRLIP-----NNTMLLYLLIOVLVYI 180
 DB 100 ---LNFKDEIFN-----ELDPTLIYCGAMLANLVSFRLRL--I 136
 QY 181 ESNVNIPLKFRKRVIVYLLFLILISF-----LSQYT----- 214
 DB 137 EKVNIISFEVHKRVYIIFISVALTLVAFYNNIYAGSIAGDGSVLKIMPLFTGTIIL 196
 QY 215 -----KQVONEIMACKQOIRNITQYSOQIESLYKDIRSEPHDNLITITST 261

Db 197 IVIVTVVINTAFNENLKVQN---QKE-QLEQLQDYVTTLSELRKMRVFRHDYVNIISLT 251

Qy 262 RGIENKDLASIEKTIYHOILEKTHQLODTRVYNGHIANIQNDAVNGILISAKIILEAONKK 321

Db 252 VGYIDNNMPGIKYYFENNVIPIKNTESNNYKISLQNNHYTELKGLAVLIRAOELK 311

Qy 322 IAVNVESSKIQLEPMELDFITLISLQDNATAEAFESINPEIOLAFPKKNGSIIVFIIO 381

Db 312 IDALEVEEPIDKISMSIDCKYVGLLDNAVEALTCENPVIKRAFAVKKGDSIIIVFA 371

Qy 382 NSTEKQIDVSKIREKNTSTGSKRGIGLAKVNHILEHPKTSIQ 427

Db 372 NSLPVNMPIYKIFEEGSGTKGEGRGILASLEIRIMKYSHVADLT 417

RESULT 7

AD1438

sensor histidine kinase (Agrc from Staphylococcus) homolog lin0043 (imported) - Listeria

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AD1438

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloechet, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AD1438

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-431 <GLA>

A:Cross-references: CA:AL597022; PIDN:CA095276.1; PID:g16412464; GSPDB:GN00178

A:Experimental source: Strain Clp11262

C:Genetics:

A:Gene: lin0043

Query Match 19.18; Score 416; DB 2; Length 431;

Best Local Similarity 26.28; Pred. No. 3.7e-18;

Matches 126; Conservative 94; Mismatches 148; Indels 112; Gaps 16;

Qy 21 LILFESKYSNNTLSKKEIT--LESISNPLMIAVTMNVNLPAP--PLYFIAL-----S 72

Db 4 IIMATIQITGIFIAQIITKVFSEIKGELVTIAMAFELFVQVMSMIFVLIVFVSA 63

Qy 73 IYLNQN-----SLSNIFYGLPVAASDLFRRALIFLIDGTGIVMGSSITTYMIEF 127

Db 64 LWMKKNIVSGSITLVYII-LFTISDS-----IVGFIL-----VPG----- 99

Qy 128 AGIALSYFLSYFVNDIGRLKDSITKMKVKRLIP-----NITMLLYLLIOVLYI 180

Db 100 ---LNFYKDELFN-----ELPLTYICAGMLANLIVFSPILKRL-I 136

Qy 181 ESYNVIPLTKRKFVYVYLLIFLILISF-----LSQYT--- 214

Db 137 EKVNISRFVEHRKAYIIFISIVALTVAFYMNIVAGSIAGDSVLKINTLLFTGYTIDL 196

Qy 215 -----KQVONEIMAKOEKQIRITQYSSQISLKYDRSRPHYINILNSL 261

Db 197 IVIVTVVINTAFNENLKVQN---QKE-QLEQLQDYVTTLSELRKMRVFRHDYVNIISLT 251

Qy 262 RGIENKDLASIEKTIYHOILEKTHQLODTRVYNGHIANIQNDAVNGILISAKIILEAONKK 321

Db 252 VGYIDNNMPGIKYYFENNVIPIKNTESNNYKISLQNNHYTELKGLAVLIRAOELK 311

Qy 322 IAVNVESSKIQLEPMELDFITLISLQDNATAEAFESINPEIOLAFPKKNGSIIVFIIO 381

Db 312 IDALEVEEPIDKISMSIDCKYVGLLDNAVEALTCENPVIKRAFAVKKGDSIIIVFA 371

Qy 382 NSTEKQIDVSKIREKNTSTGSKRGIGLAKVNHILEHPKTSIQ 427

Db 372 NSLPVNMPIYKIFEEGSGTKGEGRGILASLEIRIMKYSHVADLT 417

Db 372 NSLPVNMPIYKIFEEGSGTKGEGRGILASLEIRIMKYSHVADLTPTKTNREVIOELIEM 431

RESULT 8

agrc protein - Staphylococcus aureus

C:Species: Staphylococcus aureus

C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 15-Oct-1999

C:Accession: S58480; S20797

R:Novick, R.P.; Projan, S.J.; Korblum, J.; Ross, H.F.; Ji, G.; Kreiswirth, B.; Vande Mol, Gen. Genet. 248, 446-458, 1995

A:Title: The agr operon: an autocatalytic sensory transduction system in Staphyloc

A:Reference number: S58478; MUID:96004766

A:Accession: S58480

A:Molecule type: DNA

A:Residues: 1-423 <NOV>

A:Cross-references: EMBL:X52543

A:Note: the authors translated the initiation codon GTG for residue 1 as Val

R:Korblum, J.; Projan, S.J.; Kreiswirth, B.N.; Mogazeh, S.L.; Elsnar, W.; Ross, H.; submitted to the EMBL Data Library, March 1990

A:Reference number: S20793

A:Accession: S20797

A:Molecule type: DNA

A:Residues: 1-83, 'T', 85-423 <KOR>

A:Cross-references: EMBL:X52543; NID:946505; PIDN:CA436783.1; PID:g581546

C:Genetics:

A:Gene: agrc

A:Start codon: GTG

Query Match 17.48; Score 379; DB 2; Length 423;

Best Local Similarity 21.98; Pred. No. 6.4e-16;

Matches 96; Conservative 107; Mismatches 163; Indels 72; Gaps 7;

Qy 33 LSKKELTESISNPLMIAVTMNVNLPAPLYFIALSIYLNQNSLSNIFYGLLPV 92

Db 4 LNSYNEVLFVLQMLMTFTIPALISGIVSKDYFLI-----IYI 43

Qy 93 ASSDLFRRALIFLIDGTGIVMGSSITTYMIEFAGIAL-----SYLFLSV 139

Db 44 STLSLF---LFRKMPDSALILTSFIIMFYKIKYYSILLMSQIILYCANYWYIYI 99

Qy 140 FVNDIGRLKDSITKMKVKKRLIPNITPMLLYLLIOVLYESYVNIPLTKPKK-FVTV 198

Db 100 Y-AYTKISDSIFVIFPSEFVYVITISLFSYIINRYLKLST---PYILNKGGLIYI 154

Qy 199 YLILFL-----LISFLSYTRQVONEIMAKOEKQ 229

Db 155 STILTLTFELFFYSQINSDEKAVIROYSLFYWYHILYSILTYSQLKEMKVKRNOEE 214

Qy 230 INRITQYSSQISLKYDRSRPHYINILNSLRLGIEKNDLASIEKTIYHOILEKTHQLO 289

Db 215 IETYYTYTKIAINNEKRRHDYVNIITLSEYIKEDDMPGIRDFYFNKNYIPKMDNQ 274

Qy 290 DTRYNIGHLANIQNDVKGILSAKILEAONKKIAVNVSSKIQLEPMELDFITLISLT 349

Db 275 MNAIKLNGIENLKVREIKGLITAKILRAOENMIPISIEIDEVSSINLNMIDLSRSIGII 334

Qy 350 CDNALTAEPESINPEIOLAFPKKNGSIYFIQNTKQKQIDVSKIREKNTSTGSKRGIG 409

Db 335 LDNALTAESTEDPIIRAFIESENSYFIYMNKCADDIRIHELQVDESGTSGEGRGIG 394

Qy 410 LAKVNHILEHPKTSIQ 427

Db 395 LSTLKEIADNADNVLDDT 412

RESULT 9

D89995

accessory gene regulator C (imported) - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: D89995

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizukami, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shib, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsuka, K.
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Accession: D89955; MUID:21311952; PMID:11418146

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-371 <OR>

A:Cross-references: GB:BA000018; PID:g13701832; PIDN:BA843125.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: agrC

Query Match 17.1%; Score 372; DB 2; Length 371;

Best Local Similarity 23.5%; Pred. No. 1.5e-15;

Matches 101; Conservative 92; Mismatches 162; Indels 74; Gaps 8;

14 TYLVFLFLFLFSKVNLTLSKKELEFISNFMIAVTWVNLFF---YPAEPLVF 68

4 TFSIIFLFLFKLYVAIVT-----LVMTIMVLSNFTVGLFLTKYTTDPA-- 52

69 IASIVLRNNSLNIIEGGLPVASDLFRRAIIFPILDTGQIVMGSSITTYMTEFA 128

53 ILPLVILSSVSLATY--LVKISLKKFKSYL-----SLNKTYMIIS 96

129 GIAL-SYLVSPNDIGRLKSLTKMKVKRLIPMNITMLLYLLIOVLYESYVNP 187

97 FVLEAFAPFYIXSTWSSNGDS-----LIPALVFLGLIFISVILLMSLFLK 147

188 TLKPRKVVIVLLEFLILISFSQYTKQVONELMAKQKQINPTQSOISLKYDI 247

148 EMYKR-----NQEIEYTYTKIEINNM 175

248 RSRPHDYLTLSLGLIENKDLASIEKTYHQLEKTHQLODTRYNIGHLANTQDAVK 307

176 RKRPHDYLTLSLGLIENKDLASIEKTYHQLEKTHQLODTRYNIGHLANTQDAVK 307

308 GILSAKLEONKKAIVANVSSKIQLEPMLDFTITSLICDNALPAFESLPEIQL 367

236 GILTAKTAKROENIPISIEIPREVSSIMLMDLSRIGITLIDNALPASTEIDPIIRV 295

368 AFRRKNGSIVFIIONSTKQIDVSKIFRENTSTGNSFGIGLAKVNHILEHPTSLQT 427

236 AFTESENSVFIYMNKCADIPRIHELFOESSTGEGRLQSLTLKEIADNADVLDT 355

428 SNHHLEFKQ 436

356 IIEGPFRIQ 364

RESULT 10

G56273

protein-histidine kinase homolog sspK - *Lactobacillus sake* (strain Lb706) plasmid pLSA60

C:Species: *Lactobacillus sake*

C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 15-Oct-1999

C:Accession: G56273

R:Axelsson, L.; Holck, A.

J. Bacteriol. 177, 2125-2137, 1995

A:Title: The genes involved in production of and immunity to sakacin A, a bacteriocin fr

A:Reference number: A56273; MUID:95238285

A:Accession: G56273

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-428 <AXE>

A:Cross-references: GB:Z46867; NID:959849; PIDN:CAA86944.1; PID:g59854

C:Genetics:

A:Gene: sspK

A:Genome: Plasmid

A:Start codon: GTC

Query Match 16.7%; Score 362; DB 2; Length 428;

Best Local Similarity 24.8%; Pred. No. 7e-15;

Matches 111; Conservative 84; Mismatches 178; Indels 74; Gaps 12;

12 LTVLVFLFLFLFSKVNLTLSKKELEFISNFMIAVTWVNLFFPA 63

39 IAVFPPIPLVAVLSDISNVIPILSGVILKEKENDYLLINDLFCFTIFECNV----- 94

64 EPLVFLAISIVLRNNSLNIIEGGLPVASDLFRRAIIFPILDTGQIVMGSSITTYMTEFA 123

95 -----LSSNIMQLIP-----SNRIVGFF-----GIFQL 119

124 MIEFAGIALSYLVSPNDIGRLKSLTKMKVKRLIPMNITMLLYLLIOVLYESYVNP 180

120 FIEALVISTITFFYRKNH-----NVKEKASTVSTGLILLYLLISTYA 168

181 ESYNVIFPLKREKV--VIYVLL--FLIISFLSQYTKQVONELMAKQKQINPTQSOISLKYDI 235

169 HYDA-----YDHFVLGIMFLIQVFEVET-FLRMLTQRTKRYKQLEKQ-ELNNLKK 221

236 YSOQIESLYKDIRSRHDYLTLSLGLIENKDLASIEKTYHQLEKTHQLODTRYNIGHLANTQDAVK 307

222 YTESLEQOOQOQSKRPHDYLTLSLGLIENKDLASIEKTYHQLEKTHQLODTRYNIGHLANTQDAVK 307

296 GHLANTQDAVKGILSAKLEONKKAIVANVSSKIQLEPMLDFTITSLICDNALPAFESLPEIQL 367

282 KALYNHNEFVSLITAKIHOAKELNTECYECQKPLDIPIPFDCIRLIGLIDNATE 341

356 AAFESINPEIOLAFRRKNGSIVFIIONSTKQIDVSKIFRENTSTGNSFGIGLAKVNHILEHPTSLQT 427

342 AASECEKTLVLYIYODDLQIEFSIKNTKYSNNSISITQIRKNSITRKSGSLGTLIDE 401

416 ILEHPTSLQTSNHHLEFKQ 436

402 ENQKPPNVFTQYKQESFVSQVLLITIK 428

RESULT 11

S57909

probable histidine protein kinase sspK - *Lactobacillus sake*

C:Species: *Lactobacillus sake*

C:Date: 24-Mar-1997 #sequence_revision 24-Mar-1997 #text_change 15-Oct-1999

C:Accession: S57909

R:Huehne, K.; Holck, A.; Axelsson, L.; Kroeckel, L.

submitted to the EMBL Data Library, February 1995

A:Description: Cloning and nucleotide sequence of genes involved in regulation and se

A:Reference number: S57909

A:Accession: S57909

A:Molecule type: DNA

A:Residues: 1-448 <HUP>

A:Cross-references: EMBL:Z48542; NID:9695615; PIDN:CAA86426.1; PID:9695616

A:Experimental source: strain Lb674

C:Genetics:

A:Gene: sspK

Query Match 16.0%; Score 348; DB 2; Length 448;

Best Local Similarity 26.3%; Pred. No. 5.3e-14;

Matches 121; Conservative 92; Mismatches 199; Indels 46; Gaps 15;

8 LNSGLITVTVLFLFLFSKVNLTLSKKELEFISNFMIAVTWVNLFFPAEPLVF 67

6 VSVSLMONFVALILFLFLYRYIORKIFPKRIID-----ILAIIFSYLIPISDASIL 59

68 FIAL-----SIYLRN-----SLSNIFYGILPVASSDLFRRAIIFPILDTGQIVMGSSITTYMTEFA 113

60 VAVLKLGNHFRQCKENIKITTDANLLILVOLLVANGIISQTTISIKSDISQNI 119

114 VNGSITTYMTEFAIGIALSYLVSPNDIGRLKSLTKMK-VKKRLIPMNITMLLYL 172

120 LNSASDITLL-----GIFFAVLFDGLFPI--LTKNRKTELQRLNDLIEFSIEK-QYPI 171

OY 173 LIOVLT-VIESNVPTLK-FREVVIVYLLLELILIS-----FLSOYT-KOKVONE 221
 Db 172 FIFITVIEILLAVGNLOQVATALLITLIFCVLLGMTFQWOMFLAKYSIROEANOQ 231
 OY 222 IMAKOEAQIRNITQVSOQIESLYKDIRSFHDYDNIITSLRLGLENKMLASIKIKYHOTL 281
 Db 232 LVANQOQO-----DVLVNEQOQYFELRFRKNDYQNIITLSLESAKRGDQOQFAYVQELL 286
 OY 282 EKGHOLQDTRYNIGHLANIQNDVAKGILSKILEAOONKRIAVNVEYSKIQPEMELL 341
 Db 287 AQRPQISEIQAVIAQLODTLKNDRIGLIVLOKFLAAQAVTLKFEETPEIELATANLTL 346
 OY 342 FIFITSLCDNAIEAFESLNPETIQLEFKKNGSIVFIIONSTREKQIDVSKIFKE-NYS 400
 Db 347 VIRIIGIILDMNIEQAVQETDQVSCAFLOSGDGLIETIENTA--SQVKNLQAFSELGYS 404
 OY 401 TKSNGNGIGLAKVNHILEHYKTSIQTSNHHLEFKOLLIT 440
 Db 405 TKAGRGSTLANVODLIANKQTNLPLETOIENRKLQTLMT 444

RESULT 12

B97095

membrane associated histidine kinase-like ATPase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum 14-Sep-2001 #text_change 14-Sep-2001

C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97095

R:Moelling, J.; Breton, G.; Omechenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C10

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97095

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-452 <KOR>

A:Cross-References: GB:AE001437; PIDN:AK79549.1; PID:g15024536; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC1582

Query Match 10.3% Score 223.5; DB 2; Length 452;
 Best Local Similarity 21.6% Pred. No. 2e-06;
 Matches 102; Conservative 97; Mismatches 186; Indels 87; Gaps 18;

OY 1 MNEALMISNGILLYLVLEFLFSKVSNTLSKKELEFSSINFLMIATVNVNLF 60
 Db 1 MNE-IOIYINSFITNALIAIKYFREFELGFSKRRKIYLFVLEIIVSIIISNVSCKLG 59
 OY 61 YPAEPIYF--ALSIYLMRONSLSINIFYGLPVASDLPFRRAIIFFLIDGTQIVMGSS 118
 Db 60 FKSIIILILEIILIFLCRGSKSLVKIYSWL--IENIILISLTFLIYDPKIDHYSRA 117
 OY 119 IITTYIEFAGIALSYFLSVFNVNDIGRLKDSLTK-----MKVKKRLPMNI 165
 Db 118 FINKD-----VLKIFNLILNFRDILNLEIFLMLKLCISIIKIKNS--ELNI 163
 OY 166 TMLLYLL-----IOVLVYESYV-----IPTLKRFPVIVVLLLELILI 207
 Db 164 YKOLIVMPCISYIGISFLF---YVQVLDYTGGEYIEPFISSITCYMLPFIIGLILIS 219
 OY 208 SFLSOYTKOV-----QNEIMAKOEQIRNITQVSOQIESLYKDIRSFHDYDNIITLS 260
 Db 220 IFLAATFQKMLDGEEREKESILMOQOQNLQ--FEHKKRNQNIYKNTRGVMDINKHINC 277
 OY 261 LRLGLENKDLASIEKIQILEKTHOLOTDRYNIGHLANIQNDVAKGILSKILEAQNK 320
 Db 278 LR-----NLAYSQNI--DVLKNYLNINISERIKKLDFILITGNAVADAVINKEKYLAKNE 329
 OY 321 KIANVVEVSSKIQLE---MELDFITLISLCDNAIEAF-----SLNPEQLAFKKKN 373
 Db 330 ---GIEIYCFRIMEKAGIEPIDLSIILNNSLDNAIEACRKIKNLKRLISIKSIYKN 385

OY 374 GSIVFIIONSTREKQIDVSKIFKENYSTKGS--NRGIGLAKVNHILEHYKPT 423
 Db 386 SFELIEISNSCEE--GIKYCNNNIISTKRSKSETNHGIGISINIKAEVKKYSGT 434

RESULT 13

AB1533

two-component sensor histidine kinases homolog lln0802 [imported] - Listeria innocua

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AB1533

R:Glaser, P.; Frangou, L.; Buchrieser, C.; Amend, A.; Baguer, F.; Berche, P.; Bloec

; Domiguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feht,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kref, J.; Kun, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

Ok, C.; Schueter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AB1533

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-433 <GLA>

A:Cross-References: GB:AL592022; PIDN:CAC96034.1; PID:g16413253; GSPDB:GN00178

A:Experimental source: strain C11p1262

C:Genetics:

A:Gene: lln0802

Query Match 8.5% Score 184.5; DB 2; Length 433;
 Best Local Similarity 24.1% Pred. No. 0.00044;
 Matches 110; Conservative 79; Mismatches 155; Indels 113; Gaps 25;

OY 16 LVYVLPFLFS--KVSNTLSKKELEF--FSINFLMIATVNVNLFPAEPLIFIALS 72
 Db 17 LEVLLIFFSKYKMNPRVSKERLLIFCLASL-----VOAFDSEF-YF----- 61
 OY 73 IYLMRONSLSINIFYGLPVASDLPFRRAIIFFLIDGTQIVMGSSITTYMIEFA-GIA 131
 Db 62 -----TINV-----IVSAVLFLVSEFFERKMODILM-SIATVAVIDIGEGIA 104
 OY 132 LSVLPSVNVNDIGRLKSLTKMKVKKRLIPNITMLLYL-----LIOVLVIESYV- 185
 Db 105 L--WSLSTALGR-----QVGEFMDQPSNTMLFLFISRLISLGLVLLIFRKRIR 153
 OY 186 -----IPTLKRFPVIVY--LILFL-----ISF-----LSOYTKOK 217
 Db 154 LGNSLPITYWTFPLFTMKDILMWLILADSPEISFYWAAVIPLLPVSYILFYTRRS 213
 OY 218 YONEIMAKOEQI--RNITQVSOQIESLYKDI--RSFHDYDNIITSLRGLENKDLAS 272
 Db 214 IEMVATQVDSKVLDDEKKKYEQDLVTMKQTLQESQVRVHKLKNSPL--TYLAESGK 270
 OY 273 IEKIHQILEKTHOLOTDRYNIGHL-----ANTONDAVKGISAKILEAONKRIAVN 325
 Db 271 TDELVEVOE-----LGLSLVGLKIVADSGNITIDYIILKQALANAGVYIF 318
 OY 326 VEVSSKIQLEPEMELDFITLISLCDNAIEAFESLNP-EIOAEFKKNGSIVFIIONST 384
 Db 319 CEINVPNDI-DVPPDLSTVLGNLIDNAARLEEVONEKRTIGIKISYQVGFMLIOVANSF 377
 OY 385 KE-KQIDVSKIFKENYSTKGSNRGIGLAKVNHILEHY 420
 Db 378 DGIIVHDMKRLISREKDE--NHGIGLSIKIKIAGRY 412

RESULT 14

E69600

two-component sensor histidine kinase cIts - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: E69600

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bered
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 350, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
leeh, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
Koeber, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, I.; Scanlon,
A.; Schleich, S.; Schroeder, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Setor
deuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Toynoni, A.; Tosato, V.; Uchiyama,
T.; Whittors, P.; Wypal, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A65580; MUID:98044033
A:Accession: E69600
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-342 <KUN>
A:Cross-references: GB:299108; GB:AL009126; NID:92633055; PIDN:CAB12587.1; PID:92633082
A:Experimental source: strain 168
C:Genetics:
C:Gene: cits
C:Superfamily: two-component sensor histidine kinase; sensor histidine kinase homology

Query Match 8.0%; Score 174; DB 1; Length 542;
Best Local Similarity 24.1%; Pred. No. 0.0025;
Matches 97; Conservative 73; Mismatches 134; Indels 98; Gaps 21;
QY 35 KKELEFSSNFIIMAVTVVNVNLFYPAEPYFIALSTYLN--RONSLSINIFGLPV 92
DB 171 KKHRLSLVIAVLVL-----LGFIAVAVLAKSIRKDTL-----GLEPH 209
QY 93 ASSDLF--RAIIFILDGQIVNGSSITFTYMIFFAGIALSYLFLSVFVNDIGRLKDS 150
DB 210 EIALALYRERNAMFAIRG-----IAT--NREGV-----VTMMVSAEM-- 248
QY 151 LRFKMKVKKRLIMNTMLLYLLIOVLVIESYVNPFLKPKFVIVYLLPLILISFL 210
DB 249 ---LKLPERVHLPIDDPVPG--AGLSVLEKGEMLPMQF-----VSYNDOVFII----- 293
QY 211 SOTYKQVONE-----IMAKQEAQIRNITQYSQOIESLYKDIRSRHDYLN---IL 258
DB 294 ---NKKVWNOGQAYGVIVSFRKTELEKLLIDLETVKRYSEDLRAQTHEFSNKLAYAIL 349
QY 259 TSLRLGIEKKDLASIEKLYHQLEKTHQLODTRNIGHLANIONDAVKGLSAKILEAQ 318
DB 350 GLELEGEDVDAIDIKREY--AIQNEQHDL-----LPH--NHSQOVQAILLGLKISKAS 399
QY 319 NKKIANNV-EVSSKIQLP-EMELLDFITLSILCDNAIEAFESINPEIQLAFKKNGSI 376
DB 400 EKRVKLVIDENSLAPLPAHIGLSHLITIIINLIDNFEAVAEQSVKEV-LFFITDMGHD 458
QY 377 VFIIQNST-----KEKQIDVSKIFKENSSTKGSNKGIGLAKV 413
DB 459 IYIEVSDTQPGVPPK---IEAVFERGYSKGMRRGIGLANNV 497

A:Residues: 1-532 <STO>
A:Cross-references: GB:AP001508; GB:BA000004; NID:910172890; PIDN:BA004116.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
C:Gene: BH0397
C:Superfamily: two-component sensor histidine kinase; sensor histidine kinase homology

Query Match 7.9%; Score 171.5; DB 2; Length 532;
Best Local Similarity 24.0%; Pred. No. 0.0035;
Matches 53; Conservative 44; Mismatches 87; Indels 37; Gaps 5;
QY 218 YQNEIMAKQEAQIRNITQYSQOIESL-----YDIRSFRHDYLNITLSRLGIEKKDLA 271
DB 299 IYNDVVGAIISFRDLPEIKQQAQOLGVKLYEALQAQSHFPMNKLHVILGWKTESYD 358
QY 272 SIEKTYHQLEKTHQLODTRNIGHLANIONDAVKGLSAKILEAQNKKIANNVSSK 331
DB 359 ELNDYTHQIVNHRSTELNHV-----IKRKDSVLAQFIIGKLSVAEKHTIDVQTKSV 412
QY 332 IQLPEMELL--DFTTILSILCDNAIEAFESINPEIQLAFKKNGSIVFIQNSTREKQI 389
DB 413 IPKASSEDVMEHLITITGLNLTIDNLEAVYKA-----KKKTVLQLEYSNRLHI 461
QY 390 DY-----SKIKENYSTKGSNKGIGLAKVNHILE 418
DB 462 RVQDTGPGIPDEGQDIPKKGYSTKENRGYGLYVQSVE 502

Search completed: July 30, 2002, 15:15:20
Job time: 391 sec

RESULT 15
E83699
two-component sensor histidine kinase BH0397 [imported] - *Bacillus halodurans* (strain C-
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: E83699
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hirai
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
A:Reference number: A63650; MUID:20512582; PMID:11058132
A:Accession: E83699
A:Status: preliminary
A:Molecule type: DNA

Wed Jul 31 08:30:58 2002

us-09-833-017-4.rpx

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:20:52 ; Search time 15.86 Seconds

(Without alignments)
1076.628 Million cell updates/sec

Title: US-09-833-017-4
Perfect score: 2173
Sequence: 1 MNEALMLSLNGLLTLFLVLF.....KTSLSQTSNHHHLFKOLLITIK 441

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000.

Post-processing: Minimum Match 0%,
Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	151	6.9	543	1	DCUS_ECOLI
2	139	6.4	428	1	KINB_BACSU
3	129.5	6.0	532	1	DCUS_BACSU
4	129	5.9	880	1	YE21_ARCTO
5	128.5	5.9	552	1	DP1B_ECOLI
6	127	5.8	577	1	CHVG_RHIME
7	126.5	5.8	535	1	DCUS_BACSU
8	126.5	5.8	691	1	ICN3_LACIA
9	125.5	5.8	717	1	NTP2_MSEPV
10	125.5	5.8	938	1	VL20_HSVTJ
11	120.5	5.5	430	1	STRA_BUCAT
12	118	5.4	1024	1	RPOB_PLARA
13	117.5	5.4	214	1	SP2K_BACSU
14	117.5	5.4	388	1	YF02_AQUAE
15	116	5.3	656	1	YC26_PORPU
16	116	5.3	850	1	DP01_ANAHI
17	115.5	5.3	451	1	YG1Y_HAEIN
18	115.5	5.3	873	1	COX1_ACACA
19	114.5	5.3	600	1	NIST_LACIA
20	114.5	5.3	904	1	TIR3_HUMAN
21	113	5.2	366	1	Y976_METUA
22	112.5	5.2	382	1	Y061_MESVI
23	112.5	5.2	883	1	YH6_YEAST
24	112.5	5.2	1174	1	EX5B_BUCAT
25	112	5.2	496	1	YFHK_ECOLI
26	111.5	5.1	780	1	MUS2_BORBU
27	111	5.1	556	1	NU2M_PODAN
28	110.5	5.1	886	1	RA50_ARCTU
29	109.5	5.0	440	1	CAPE_STANU
30	109.5	5.0	1274	1	BXE_CLOBO
31	109	5.0	1121	1	ALP1_SCHPO
32	108.5	5.0	455	1	TRME_LACIA
33	108.5	5.0	614	1	YDNK_LACIC

34	108	5.0	339	1	Y83C_METUA
35	108	5.0	608	1	ATOS_ECOLI
36	108	5.0	885	1	YDGH_BACSU
37	108	5.0	1036	1	YG35_YEAST
38	108	5.0	1738	1	YCF1_EPIVI
39	107.5	4.9	336	1	YPH1_MYCCA
40	107.5	4.9	490	1	PIT_BUCAT
41	107.5	4.9	2136	1	YCF2_MARPO
42	107	4.9	720	1	STFB_PORPU
43	107	4.9	844	1	BRO1_YEAST
44	107	4.9	1294	1	YA3B_SCHPO
45	106.5	4.9	660	1	Y390_MYCPN

ALIGNMENTS

RESULT 1	DCUS_ECOLI	STANDARD;	PRT: 543 AA.
ID	DCUS_ECOLI	P39272; P76795;	
AC	01-FEB-1995 (Rel. 31, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Sensor Protein dcus (EC 2.7.3.-)		
GN	DCUS OR B4125 OR Z5727 OR ECS5107.		
OS	Escherichia coli, and		
OC	Escherichia coli O157:H7.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Escherichia.		
OX	NCBI_TaxID=562, 83334;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K12 / MG1655;		
RX	MEDLINE=95334362; PubMed=7610040;		
RA	Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,		
RA	Blattner F.R.;		
RT	*Analysis of the Escherichia coli genome VI: DNA sequence of the		
RL	region from 92.8 through 100 minutes.*;		
RL	Nucleic Acids Res. 23:2105-2119(1995).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / EDL933 / ATCC 700927;		
RX	MEDLINE=21074935; PubMed=11206551;		
RA	Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,		
RA	Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,		
RA	Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,		
RA	Grobeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamousis K.,		
RA	Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,		
RA	Welch R.A., Blattner F.R.;		
RT	*Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.*;		
RL	Nature 409:529-533(2001).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=O157:H7 / RIMD 0509952;		
RX	MEDLINE=21156231; PubMed=9765796;		
RA	Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,		
RA	Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,		
RA	Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,		
RA	Kuhara S., Shiba T., Hattori M., Shinagawa H.;		
RT	*Complete genome sequence of enterohaemorrhagic Escherichia coli		
RL	O157:H7 and genomic comparison with a laboratory strain K-12.*;		
RL	DNA Res. 8:11-22(2001).		
RN	[4]		
RP	CHARACTERIZATION.		
RC	STRAIN=K12 / MC4100;		
RX	MEDLINE=9840434; PubMed=9765574;		
RA	Zientz E., Bongaearts J., Unden G.;		
RT	*Fumarate regulation of gene expression in Escherichia coli by the		
RL	dcus (dcus genes) two-component regulatory system.*;		
RL	J. Bacteriol. 180:5421-5425(1998).		
RN	[5]		
RP	CHARACTERIZATION, AND TOPOLOGY.		

Query Match	6.9%	Score 151	DB 1	length 543
Best Local Similarity	23.4%	Pred. No. 0.041		
Matches	89	Conservative	68	Mismatches 143; Indels 80; Gaps 18;
QY	85	IFYGLTPVASSDLEF--RAAIIFFILDCGTQGIVMSSIIITMYIEFAGIALSYL-----F	136	
Db	208	ILFGLFLEYELSTLFEDRQAMLSIQEGVAVADDSKV--FLINAAQELLNTRKSDDEK	265	
QY	137	LSVFN-----VDIGR-LKDSLTK---MKYKRRIIPNITMLLYLLIOVLVYIESNV	185	
Db	266	LSLTSHMSQVADVSEVLDRDGTPRDEELTIKDRLLIN-----	304	
QY	186	IPPLAKPKFVYIYLLILFLILSPISQYTKKQVQNEIMAKQEAQIRNITQYSSQIESLTK	245	
Db	305	--IVPVKSNCVI-----IGALSTF-----RKTEVRLKMDRGDLGNVAD	342	

[illegible]

```
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```

QY 127 --FAG--IALSYLFLSVFNDIGRLKDSLTK----- 153

```

Db 187 PGTWGSLLASHIKRQTFNMPEDLAHLVLRDASPAHIEGVAINKHETITMEAR 246
Oy 154 --MKVKKRLIPMNITMLLYLLIOVLYISVNIPTLKRFRFVIT--VYLLIFLI--- 205
Db 247 RMLGVKRAIGRNI-----HEVTPDKLPPELLISIGKPLPYOREFYIOGR 289
Oy 206 LISFISQYTKQVQNEIWA-----OKEAOIRNIQYQSQIESYKIDRSFRHYLNLIS 260
Db 290 LV--FSNKPILQIDGELYGAIAIFODKSDVDRLAEELTGVAEPDALRVONHESKHLH 347
Oy 261 ---LRIGIENKDLASIEKTYHQLEKTHQLODTRYNIGHLANIONDAVKGLSKALE 316
Db 348 IAGLLQDEGKKALQYIPDLLEEBOEFSGVYMQ-----KIHNDSLAGLLGKVS 397
Oy 317 AQNKRIANVVEYSK--IQLPK--MELLPFTITSLICDNALIEA--AFESLNPEIQALAFKK 372
Db 398 GKELQVOYIIEKDSFIDHPEGVTTHDLVYVGNLIDNSLDAPSIOQDKKTYHVFIGE 457
Oy 373 NGSIVFITQNSTREKQIDV-SKIFKENSYTKG--SNRGIGLAKVNHILE 418
Db 458 NDFLKIRVNDGEGIREYREKMEYRGFSTKSTSGRGIGLFOAIVE 505

RESULT 4
YE21_ARCFU STANDARD: PRT: 880 AA.
ID YE21_ARCFU
AC O28851.
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DB Hypothetical protein AF1421.
GN AF1421.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota: Archaeoglobales: Archaeoglobaceae:
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrle N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftis B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0182 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AE001005; AAB89825.1.
DR TIGR: AF1421.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 19 41 POTENTIAL.
FT TRANSMEM 61 83 POTENTIAL.
FT TRANSMEM 103 125 POTENTIAL.
FT TRANSMEM 152 174 POTENTIAL.

```

```

FT TRANSMEM 195 217 POTENTIAL.
FT TRANSMEM 237 259 POTENTIAL.
FT TRANSMEM 268 290 POTENTIAL.
SO SEQUENCE 880 AA; 102210 MW; 8AD99AB7B4E50671 CRC64;

Query Match
Best local similarity 20.7%; Pred. No. 1.4;
Matches 113; Conservative 82; Mismatches 177; Indels 174; Gaps 27;

Oy 10 NGLLTYLT-VLEFLFESKYSNNTLSKELLTFLSISNELLIMAVVNVNLTPEAPL-- 66
Db 52 NVFLTYLKHSGFFFL-----EPITFVPLFLTNAAVR--KVTLRHSGBPLKI 97
Oy 67 -----YFALSIVLNQNSLSNIFGLDPAVSDIFRRALIFLLDGT----- 110
Db 98 PHLADFGIAFFAALTIV--KNWSSMLFY-----FNSSNFQVSDPIFGIDAAFYTFOLP 148
Oy 111 -QGIYMGSSITTYMIEFAGALSYFLSVFNVDIQLADSLTKMKVKKRLIP---MNI 165
Db 149 FLKIVIG-SLTAALLISLALAAFAVMTAF-----RWKSLDEE--KEIPGSGFMHF 197
Oy 166 TMLLY--YLLIOVLYIESYVNI-----PTLKE----- 191
Db 198 SALLFASPLLSAALYLRSREIVHSEHGLISGASYVDVNLISLMLFAIVLLSGIFAA 257
Oy 192 -----RKVVIVVLLLELLISFLSOYTKOK--YONEIWAQKEADIRNITQSOOI 240
Db 258 YIVARRVERVQIVGVLFVAILTLTVAPPVQKVEYEPSEISYGRYIANSINTLFA 317
Oy 241 ESLYKDIRSPHDYLNLTLSRLGLENKDLAS-----IEKIQHOLE-RTGHQLOD 290
Db 318 YNLH-DVKLQKFDYVDVYVNDI-LEAKPTIDNRIMDHPRIDVPROLQOIRTYVIOD 375
Oy 291 T---RYNIGHLANIONDA-VKGLISAKILEAONK-----LVANVE 327
Db 376 VDVKRYI-----NDISYQULAAARELSTENLPNSAQTWLNKLLTYRGCIYASPN 428
Oy 328 VSKTIQPEMELDLFTITLSLICDNALIEAFAESLNPEIQALAFKKNGSIVFITQNSTREK 387
Db 429 KYSKRGLEDEFYIYDI-----PPEGLIKIERPEIYVGGELTDYVVVNTLQDEFFDP 478
Oy 388 QIDVSKIRKENYKSK--GSGNGIGL-----AKVNHILEHYPKTSLOSGNHHHLF 434
Db 479 KGDV-----NFTKIVAGDGCIKLDYRKVLFSPRGEDINLISNITTESRLAMHROIV 532
Oy 435 KQLLTI 440
Db 533 ERVSTI 538

RESULT 5
DP1B_ECOLI STANDARD: PRT: 552 AA.
ID DP1B_ECOLI
AC P77510; O54338;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Sensor kinase dp1b (EC 2.7.3.-) (Sensor kinase cta).
GN DP1B OR CTA OR MPDB OR B0619.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=KL2 / MG1655;
RX MEDLINE=98367124; PubMed=9701802;
RA Ingner H., Miller C.A., Cohen S.N.;
RT "Destabilized inheritance of pSC101 and other Escherichia coli
RL plasmids by DplA, a novel two-component system regulator.";
RN Mol. Microbiol. 29:49-59(1998).

```

RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9728503;
 RA Blatner F.R., Plunkett G., Tarr, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
 RA Nederspiel N., Hyman R., Kaiman S., Komp C., Kuri O., Lew H., Lin D.,
 RA Natch A., Oetner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97061202; PubMed-8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sasaki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiiuchi T.;
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DP1A/DP1B
 ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES
 AND GENES INVOLVED IN PLASMID INHERITANCE. MAY ACTIVATE DP1A BY
 PHOSPHORYLATION.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 (Potential).
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U46667; AAC28951.1; -
 DR EMBL: AE000167; AAC73720.1; -
 DR EMBL: U82598; AAB40819.1; -
 DR EMBL: D90702; BAA35255.1; -
 DR EMBL: D90703; BAA35261.1; -
 DR Ecogene: E613646; dplB.
 DR InterPro: IPR004358; BCRRLSENSOR.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_S19.
 DR InterPro: IPR000014; PAS.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF00989; PAS; 1.
 DR PRINTS: PR00344; BCRRLSENSOR.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00091; PAS; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; FALSE-NEG.
 KW Sensory transduction; Transferase; Kinase; Phosphorylation;
 KW Transmembrane; Inner membrane; Complete proteome.
 FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 22 182 POTENTIAL.
 FT DOMAIN 43 482 PERPLASMIC (POTENTIAL).
 FT TRANSMEM 183 203 POTENTIAL.
 FT DOMAIN 204 552 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 222 292 PAS.
 FT DOMAIN 344 541 HISTIDINE KINASE.
 FT MOD_RES 347 347 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT CONFLICT 518 552 GGVTLLEDNPGCLTFSTIYPRKPKNDSSINPIDR -> VV
 FT SEQUENCE 552 AA: 61684 MW: 3535081ABD6E77 CRC64;
 SQ SEQUENCE 552 AA: 61684 MW: 3535081ABD6E77 CRC64;
 Query Match
 Best Local Similarity 21.5%; Score 128.5; DB 1; Length 552;
 Matches 90; Conservative 81; Mismatches 150; Indels 97; Gaps 20;
 OY 18 VPELFLFSKYVNTYLSKELTFLSINFLIMVWNNLPFA-----EPLIF 68
 DB 162 VVSTIVLSKIDIS--WAEFLPAGYVVLGLIML-LSWFLAIIIRQMGMEK-- 215
 OY 69 IALSTYLNKNSLSLNIFFGLPYASSDIFERRAIFFLDGTGIWSSIIITYMIEFA 128
 DB 216 -QIARVVRQGEALFSSVYEGILAV-----DPHYITAINRNARKMLGLS 258
 OY 129 GIALSYLPLSVNNDIGLKSILTKMKYKRLIMNTMLLYLLIOVLYIESVYIP 188
 DB 259 SPGRKWLGRPI--VEVVRPADFTFQIDEKR-----QDVANFGLSV 299
 OY 189 LKFRFVVIYVLIIFLILISFLSQYTKOKVNEIMAKREAIRNTOYSQIESLYKDIR 248
 DB 300 IANREAINSGDGLGAL-ISFRSK-----DEISLNNQIOLIKY--VESL---R 343
 OY 249 SRRHVYLIILSLRGITENKDLASTIEKTYHOILEKTHQLODTRYNGHLANIQND-AYK 307
 DB 344 YLRHEHLMWSTLNLQMKR-----YDRVLAIVGEGESAOQDLISLREAFAROVA 396
 OY 308 GILSAKILEAOKKIAVAVESSKI-OLPE-MELLDEYTLISLDCDAIAPAES--LNP 363
 DB 397 GLLPKQVQARERELGKMTIYVGSQLSQPLPGLDSTERRAIVGNLDNAFASLRSDGK 456
 OY 364 ETQLAFEFKKNISYFI-----IQNSTREKQIDVSKIRENTSTKG---SNRGIGL 410
 DB 457 IVEL-FLSDEGDVYIEVADGCGVPSLRD-----KIFEGVSTRADPEGRHGIGL 507
 RESULT 6
 CHGV_RHIME STANDARD; PRT; 577 AA.
 ID CHGV_RHIME STANDARD; PRT; 577 AA.
 AC P72292; 031069;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 OS Sensor protein chvg (EC 2.7.3.-) (Histidine kinase sensory protein
 exos).
 DE CHGV OR EXOS OR R00043 OR SMC04446.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 CC NCBI_TaxID=382;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-98083050; PubMed-9422587;
 RA Cheng H.-P., Walker G.C.;
 RT "Succinoglycan production by *Rhizobium meliloti* is regulated through
 the ExoS-chv1 two-component regulatory system.";
 RL J. Bacteriol. 180:20-26(1998).
 RN (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN-1021;
 RX MEDLINE-21396507; PubMed-11481430;
 RA Capela D., Bariloy-Hubler F., Gouzy J., Bothe G., Anpe F., Batut J.,
 RA Bolstad F., Becker A., Bouly M., Cadieu E., Dreano S., Gloux S.,
 RA Godtard T., Goffeau A., Kahn D., Kiss E., Lelaur V., Masny D.,
 RA Pohl T., Portetelle D., Puenher A., Punelle B., Ramsperger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 RN [3]

RP SEQUENCE OF 1-112 FROM N.A.
 RX STRAIN-Su47 / 1021;
 RX MEDLINE=96032383; PubMed=7559334;
 RT Oesterhaus M., Stanley J., Finan T.M.;
 RT "Identification of Rhizobium-specific intergenic mosaic elements
 RT within an essential two-component regulatory system of Rhizobium
 RT species.";
 RL J. Bacteriol. 177:5485-5494(1995).
 CC -1- FUNCTION: MEMBER OF A TWO-COMPONENT REGULATORY SYSTEM
 CC CHRG(EXOS)/CHVT INVOLVED IN REGULATING THE PRODUCTION OF
 CC SUCCINOGLUCAN. ACTIVATES CHVT BY PHOSPHORYLATION.
 CC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF027298; AAB96631.1;
 DR EMBL: AL591782; CAC1430.1; ALT_INIT.
 DR EMBL: U32941; AAB07686.2;
 DR InterPro: IPR004358; ECTRSENSOR.
 DR InterPro: IPR003660; HAMP.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR003661; HIS_KIN.
 DR Pfam: PF00672; HAMP.C; 1.
 DR Pfam: PF02518; HAMP.C; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PR00344; ECTRSENSOR.
 DR SMART: SM00304; HAMP; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00388; HASKA; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR Sensory transduction; Transferrase; Kinase; Phosphorylation;
 KW Transmembrane; Inner membrane; Exopolysaccharide synthesis;
 KW Complete proteome.
 FT DOMAIN 1 29 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 30 50 POTENTIAL.
 FT DOMAIN 51 260 PERIPLASMIC (POTENTIAL).
 FT TRANSMEM 261 281 POTENTIAL.
 FT DOMAIN 282 577 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 347 575 HISTIDINE KINASE.
 FT MOD.RES 350 350 PHOSPHORYLATION (BY SIMILARITY).
 FT CONFLICT 136 136 L -> F (IN REF. 1).
 FT CONFLICT 266 266 V -> A (IN REF. 1).
 SQ SEQUENCE 577 AA; 63720 MW; BEFD2DA4AF0E2B1S CRG64;
 Query Match 5.8%; Score 127; DB 1; Length 577;
 Best Local Similarity 22.5%; Pred. No. 1.1;
 Matches 68; Conservative 57; Mismatches 89; Indels 88; Gaps 14;
 OY 197 IYLLIFLLIFLSQYRKQ-----KVQNTI--MAKQKQRIITQYSCQIE 241
 DB 272 LVNVLISLLSSTIANPLRLAAAIIRVRGGAEREIRIPFSSRODEITGLSVALLREMT 331
 OY 242 STYKDD--IRSF-----RHDTYINLTSLRLGIENKDLASIKIYHQLLEKTHGLOQTRY 293
 DB 332 TALYDRIATIEPNADVSHELKNPLISRAVEITPLIARNESESKRLMDVYQHVR----- 367
 OY 294 NIGHLANIONDAVKGLSKILAEONKKIAYN-----VEVSKIQ-LPEMELLDFIT-- 344
 DB 388 RLDRLISDISDASR--LDAELARADAKRVDEKLGDLVEISROIROSKKPVLLDFVDR 445
 OY 345 -----ILSILCDNIEAFAESLNFELIQLAFKKNG----- 374

DB 446 KDNPRASFVSGYELRIGQITNLNLENA-----RSFVPE-----QNGRIYVRLRSLR 493
 OY 375 -STVIIONSTREKQIDVSKTFKFNSTK-----GSRNGIOLAKNHHLEHPKSLQ 427
 DB 494 RIIVYEDNGFGIQAEDRIDRIFERYDREGDEFGQNSGLGISRQIAEAHGT-LNA 552
 OY 428 SN 429
 DB 553 EN 554
 RESULT 7
 ID DCTS_BACSU STANDARD; PRT; 535 AA.
 AC P96601;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Probable C4-dicarboxylate sensor kinase (EC 2.7.3.-).
 GN DCTS.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 OX NCBI_Taxid=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=96000887; PubMed=9341680;
 RA Bejloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,
 RA Kasahara Y., Alonso J.C., Le Hegarat F.;
 RT "Characterization of an lrp-like (lyrC) gene from Bacillus subtilis.";
 RL Mol. Gen. Genet. 256:63-71(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION, AND GENE NAME.
 RC STRAIN-168;
 RX MEDLINE=20170658; PubMed=10708364;
 RA Asai R., Baik S.-H., Kasahara Y., Moriya S., Ogasawara N.;
 RT "Regulation of the transport system for C4-dicarboxylic acids in
 RT Bacillus subtilis.";
 RL Microbiology 146:263-271(2000).
 CC -1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR.
 CC PROBABLY ACTIVATES DCTR BY PHOSPHORYLATION (BY SIMILARITY).
 CC ESSENTIAL FOR EXPRESSION OF DCTR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AB001488; BAA19282.1;
 DR EMBL: Z99106; CAB12252.1;
 DR Subtilist; BG12073; dcts.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR000014; PAS.
 DR Pfam: PF00518; HATPase_C; 1.
 DR Pfam: PF00512; signal; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00921; PAS; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; FALSE_NEG.
 KW Sensory transduction; Transferrase; Kinase; Transmembrane;

Phosphorylation; Complete proteome.

FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 12 32 POTENTIAL.

FT DOMAIN 33 172 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 173 193 POTENTIAL.

FT DOMAIN 194 535 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 213 276 PAS.

FT DOMAIN 333 528 HISTIDINE KINASE.

FT MOD.RES 336 336 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

SQ SEQUENCE 535 AA; 59942 MW; 614BDAB8BF10CPE CRC64;

Query Match
Best Local Similarity 18.7%; Score 126.5; DB 1; Length 535;
Matches 84; Conservative 91; Mismatches 137; Indels 137; Gaps 24;

63 ADELIALALSTLNQNSLSLNFYGLLPVASSDLFRALIFLLDGTQGVMSSTITT 122
121 ADELIALSE---AKGEIGTAVRAFY---PVKDQDLNQG---VLVQKTL--- 160
123 YMEFAGIA--LSYL---FLSVFNVDIGRLKSLTKMKYKRLIPMNTMLLYLLIQ 175
161 ---PGIADILHLKRDIAFIVLTLGFGLAGSFLARHIKQ--- 199
176 VLVIESYVVI---PLAKFKFVVIYVLLFLFLISFLSQYTKQ---KVONEIMAO--- 225
200 -MQLPHEIVMRYEERATFMSNEGVAIDNRLVITTEFNKAKQFEVQGLIGKVIW 238
226 ---KEAOI---RNITQSOQIE---SLYKD--- 246
259 EVLKDSRLPEIYVRNKAVNEIRVSKVIMSSRIPIYMKKVIQAVAFEDPTEAKMA 318
247 ---RSF---RHLYNLITS---RLGIENKDLASIEKIYHQLKRGHLOD 290
319 EELTGKRVNVEALRVONHEHMKHTIAGLIOLGKSEKAL---QLAFDAST 366
291 TRYNIHL--ANIONDAVKGLISAQILEAONKKIAVANVESSKI-QDPE-MELLDFITL 346
367 EENVTVEFLHRSIONDAAGLLSLKIRSGRELGAIVHDESSSQDFEVDQHDIVLL 426
347 SIICDNVAIAFEESLNP---IQLAFPKKNGSIYVIO--NSTKKEQIDVSKIFEENYS-T 401
427 GNLLENAF-GSEFTYVQSEKRIIDISIEQTDLLALILEDNCGGLEPHMPALYDKGTVN 465
402 KGSNKGIGLAKVNHILEHYPKTSIQTSNH 430
486 KTGGTGYGLYVKQIIDKSGT-IEVDSH 513

RESULT 8
LCN3_LACLA STANDARD; PRT; 691 AA.

AC P37608;
DT 01-OCT-1994 (Rel. 30; Created)
DT 01-OCT-1994 (Rel. 30; Last sequence update)
DT 01-MAR-2002 (Rel. 41; Last annotation update)
DE Lactacin 481/lactococcin transport ATP-binding protein LcnD3.
GN LCNDR3.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillales/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-ADRIA 85LO30;
RX MEDLINE=94288641; PubMed=8017945;
RA Rince A., Dufour A., le Pogam S., Thuaud D., Bourgeois C.M.,
penec J.P.;
RT Cloning, expression, and nucleotide sequence of genes involved in
production of lactococcin DR, a bacteriocin from Lactococcus lactis
subsp. lactis.";
RT Appl. Environ. Microbiol. 60:1652-1657(1994).
CC -1- FUNCTION: PROBABLY IMPLICATED IN THE EXPORT PROCESS OF THE

LANTIBIOTIC LACTICIN 481/LACTOCOCCIN DR.

-1- SUBCELLULAR LOCATION: Integral membrane protein.

-1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. HLYB SUBFAMILY.

THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sdb.ch/announce/> or send an email to license@isb-sdb.ch).

CC EMBL: U91581; AAC72259.1; -
CC InterPro: IPR003593; AAA.
CC InterPro: IPR001140; ABC transporter_tmem.
CC InterPro: IPR003439; ABC transporter.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00664; ABC_membrane; 1.
CC SMART: SM00382; AAA; 1.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC Transports; Protein transport; Bacteriocin transport; Transmembrane;
CC ATP-binding.
CC TRANSMEM 157 179 POTENTIAL.
CC TRANSMEM 189 208 POTENTIAL.
CC TRANSMEM 262 284 POTENTIAL.
CC TRANSMEM 289 311 POTENTIAL.
CC TRANSMEM 381 403 POTENTIAL.
CC NP_BIND 497 504 ATP (POTENTIAL).
CC SEQUENCE 691 AA; 79834 MW; 4276DD778AEC0B47 CRC64;

Query Match
Best Local Similarity 19.5%; Score 126.5; DB 1; Length 691;
Matches 103; Conservative 96; Mismatches 179; Indels 149; Gaps 25;

14 TYLVYLFLLFL-----FSKVSNTLSKKELTFLSISNPLIMAVYV----- 55
158 TPL-YIFSLFISQIALVFSLIDILNKSHDIYS---PIMMSHLYVQTLSLMKLGA 213
56 ---NVLFPAEELVFALISILNQNLSLNFYGLLPVASSDLFRALIFLLDGTQGI 113
214 QKNTMLY-----ESKISROIENG-----IFSRLYLF-RNNSVGT 248
114 VNGSSITTYMIEFAGIALSYLFSVFNVDIGRLKSLTKMKYKRLIPMNTMLLYLL 173
249 IEKINLRGTARD-GILTK-IFPSLNF-----FVFIYIYLGFTISFLTLFVI 297
174 IOVLVIESYVVIPLKRRKRVYVYLLFLFLISFLSQYTKQKVONEIMAOKEAOIRNI 233
298 KMLLYWTFESFL---SIKROANIOTQOTIDFTSVQEDLNOIDKAOANEKCVKRW 354
234 TOYSOQIESLYKDIRSF-----RHLYNLITSRLGI--ENKDLASIEKIYHQLIE 282
355 TKRSQITTFYFNKILINDIGITSAPNGCFNYICVILAMIFGIYVINOGLVSIPL---ITF 411
283 KTG-----HOLQDFYVNIHGLA-----NIONDAVKGLISAKI-- 314
412 GSGISLPSAVNOIDVAFELISRTYGNKTSIDLLENPORINDIEKHNAILIIDLISY 471
315 -LEAON---KRIAVNEVSKITQLEMELLDFTIISILC-----DNAIEAFESINPEI 365
472 SYELNNYIFNNINFSIKKGEKIAIVGKSGSKSTYFILLGLISYREYVYGGYENRQIT 531
366 QLAFFKKN---GSIVFIIONSTKKEQIDV-----SKIKENYSTK 402
532 GYVSQNMMLRKSGSLIENIVSNNSSELDICKINDVAKDVMLELYDSLPKITSQLE-N 590
403 GSNKGIG-----LAK--VNH---ILEHYPKTSIQTSNHHHFLKQL 438
591 GNLISGQIQRLILAKSLNNKKEIFWDEPSSLDNQNRIHIYKNYL 637

```

RESULT 9
MSEPV
ID NTP2_MSEPV STANDARD: PRT: 717 AA.
AC Q9YK06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoside triphosphatase II (EC 3.6.1.15) (NTPase II) (Nucleoside
triphosphate phosphohydrolase II) (NPH II).
GN MSY086;
OS Melanoplus sanguinipes entomopoxvirus (MSEPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OX NCBI_TaxID=83191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tucon;
RX MEDLINE=99102612; PubMed=9847359;
RT "The genome of Melanoplus sanguinipes entomopoxvirus."
RL J. Virol. 73:533-552(1999).
CC -1- FUNCTION: ESSENTIAL FOR VIRAL REPLICATION. PLAYS AN IMPORTANT ROLE
CC DURING TRANSCRIPTION OF EARLY MRNAs, PRESUMABLY BY PREVENTING R-
CC LOOP FORMATION BEHIND THE ELONGATING RNA POLYMERASE. ACTS AS NTP-
CC DEPENDENT HELICASE THAT CATALYZES UNIDIRECTIONAL UNWINDING OF
CC 3' TAILED DUPLICATION MRNAs. MIGHT ALSO PLAY A ROLE IN THE EXPORT OF
CC NEWLY SYNTHESIZED MRNA CHAINS OUT OF THE CORE INTO THE CYTOSOL.
CC REQUIRED FOR PROPAGATION OF VIRAL PARTICLES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAH
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF063866; AAC97810.1; -.
CC DR InterPro: IPR001410; DEAD.
CC DR InterPro: IPR002464; DEAH_ATP_helicase.
CC DR InterPro: IPR001650; Helicase_C.
CC DR Pfam: PF00271; Helicase_C_1.
CC DR SMART: SM00490; HELIC_C_1.
CC DR PROSITE: PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
CC DR ATP-binding; Helicase; Hydrolyase; Transcription.
CC FT DOMAIN 328 533
CC NP_BIND 206 213
CC SITE 331 334 DEAH BOX
CC FT SITE 331 334
CC FT SEQUENCE 717 AA: 84725 MW: 88704661 FLEA863 CRC64;

```

Query Match 5.88; Score 125.5; DB 1; Length 717;
 Best Local Similarity 18.33; Pred. No. 1.8;
 Matches 107; Conservative 97; Mismatches 172; Indels 209; Gaps 24;

```

QY 15 YLTVLFLFLFESKSNVNLTKSKELFLPSISNFMIAVWNVNLPAPAEFLFALSLSY 74
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  26 YVYMAVLF---PNNATFNSYITKEVEFYPMQFAL-----YPVYKLYMHNNINIC 75
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 75 LNR-----QNSLSINIFYGLPLVASSDLFRRATIF-----FILDG----- 109
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 76 LNRRLIYLSNEKKNISINTVYNL-----LYNNELKPEDDNIILNGKMLKISYASYS 128
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 110 ---TGGIYWGSSITTYTMI-----EPAGTALSYPLSVYNDIGLAKSLIKMKYKR 159
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 129 VTIIISQITINISLNRKYIYGIIESANLYGILSSYKONKRYFDKVL-----FSFKSELRST 184
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 160 LIPMNITMLLYL-----LIQVLY-----VIESY----- 183
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 185 MLDVQLKIFELFISKKNCIISGCTGCTGTYIVPLKFMWFMNLPFGYEFWMVSNKNKIND 244
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

```

QY 184 -----NVPTLTKPKFK-----VVIY----- 199
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 245 FIFKPNFEKNTILSLPRKALIRMGNTIKSLGFTDNISGSPITLLYKRYKKEKEYNN 304
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 200 -LILFLILISFLS---OYTKOKVONEI-----MAQEAQIRNITQYSQ 238
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 305 PLVPEVLSVAKRITINNIKHSNSYITDEIHEHDFGDIAIARFKKKRYNIRNIVLSA 364
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 239 QIESLYVDIRSEFHDYINILISLRIGI-----ENKDLASIEKIYHQLEKTRGQL 288
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 365 TTES---DIDIRIRYKNIYIYIPGVSLEPVEKTECEDKOVSLIKNTMPVSGSVIIF 421
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 289 QDTRNRYGNLANIONDAVKGLISAKILEAO--NKIIAVNVESSKIQDPEMELDEFTIL 346
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 422 YETIKKI---NEKKEILIESILIDIKYIKYITIKSKITINAINIKQNDKKH-----IH 471
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 347 SILGDMTEAFESLNPEDL---AFKK--NGSIVFIIONSTKEKOLDVSKIFRENTS 400
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 472 IILSTINIESITTTNNTVINDGKMKYOKKFLGSGTMYITESKYIORKRGVRISKGY- 530
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 401 TKGSRGIGLAKVNHLEHYPKTSLOTS---NHHLFQQLLIK 441
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 531 -----IRYSKDLQTFTEKHINYQYLMETIIVEK 559
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```

RESULT 10
 ID V120_HSV7J STANDARD: PRT: 938 AA.
 AC P52438;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Capsid assembly protein U30.
 GN U30.
 OS Human herpesvirus (type 7 / strain J1) (HHV7).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Simplexvirus.
 OX NCBI_TaxID=57278;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nicholas J.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DDA databases.
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
 CC EBV-1 23, EBV BOLF1, VZV 21, HSV-1 63, AND HCMV UL47.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U43400; AAC54692.1; -.
 CC DR Capsid assembly.
 CC FT SEQUENCE 938 AA: 110170 MW: 94839A2BF0D32BC9 CRC64;

Query Match 5.88; Score 125.5; DB 1; Length 938;
 Best Local Similarity 20.38; Pred. No. 2.5; 112; Indels 179; Gaps 22;
 Matches 92; Conservative 70; Mismatches 112;

```

QY 2 NEALMIL-----SNGILTYTVLFLFLFSKSNVNLTKSKELFLPSI----- 43
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 577 NRRLLIILFACKLMPKSNVYLSH--YLLLLHART-----LQIFVVDGSHSIIHAIQTKIF 630
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 44 -----SNFLIMIAVTMVNVNLFYPAEPLFYFALSLIYNR--QNSLSINIFY 87
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  DB 631 DNINSLQTIPIPTNPLVSLILF-----AYVHMQTYVNPVIOQTISENI-- 676
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
  QY 88 GLPVAASDLERRAIFIFLDGTGGIYWGSSITTYTMIIEFAGTALSYLE-----LS 138
  | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |

```


[illegible]

```

RESULT 12
RPOB_PLAFA STANDARD; PRT: 1024 AA.
ID RPOB_PLAFA
P21421;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6).
GN RPOB.
OS Plasmodium falciparum.
OC Eukaryotes; Alveolates; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=5835;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BW(C10);
RX MEDLINE=95107345; PubMed=7808472;
RA Gardner M.J., Goldman N., Barnett P., Moore P.W., Rangachari K.,
RA Strath M., Whyte A., Williamson D.H., Wilson R.J.M.;
RT "Phylogenetic analysis of the rpoB gene from the plastid-like DNA of
RL Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 66:221-231(1994).
RN [2]
RP SEQUENCE OF 328-1024 FROM N.A.
RX MEDLINE=91187055; PubMed=2011147;
RA Gardner M.J., Williamson D.H., Wilson R.J.M.;
RT "A circular DNA in malaria parasites encodes an RNA polymerase like
RT that of prokaryotes and chloroplasts.";
RL Mol. Biochem. Parasitol. 44:115-124(1991).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](n).
CC -1- MISCELLANEOUS: THIS RNA POLYMERASE IS ENCODED ON A CIRCULAR DNA.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)

```

DR EMBL: X75544: CAA53232.1: -
DR EMBL: X52177: CAA36427.1: -
DR EMBL: X95275: CAA64572.1: -
DR PIR: S10438: RNZOB.
DR InterPro: IPR001572: RNA_pol_B.
DR Pfam: PF00562: RNA_pol_B.1.

FT TRANSMEM 129 151 POTENTIAL.
 FT TRANSMEM 158 175 POTENTIAL.
 FT TRANSMEM 179 196 POTENTIAL.
 FT TRANSMEM 203 225 POTENTIAL.
 FT TRANSMEM 304 326 POTENTIAL.
 FT TRANSMEM 347 369 POTENTIAL.
 SQ SEQUENCE 388 AA, 45529 MW, F00E78EEB1B9941 CRC64;

Query Match 5.4%; Score 117.5; DB 1; Length 388;
 Best Local Similarity 22.4%; Pred. No. 2.6;
 Matches 80; Conservative 55; Mismatches 127; Indels 95; Gaps 15;

QY 2 NEALM-----ILNSGLITLYLVLEFLFESKVSNTLSKELT-----LFSINFLIMI 50
 DB 7 NEVLALVYSAFLISLEGEVYIILLVYKIKERKIKOSLPGLLSLS----- 60
 QY 51 AVTWNVNLEFP-----AEPLVFTALSIVLNRON-----SLSLNIFYGILLPVAS 94
 DB 61 --TVLSTAFYPRFKLGIEEGFQFIYFLNLKKEVGFSGKIFPKLLIGLILPVF 118
 QY 95 SDFERRAIFPIDGT--OGIVGSSITTYMI-----EPAGIALSYLFISVFNVDIGRL 147
 DB 119 YKFKYKGEPRFMGTFFVGFPAALFSTITFLFEKRRFTIYPLFLFLAV----- 170
 QY 148 KSLFTKAKYKRLIPNITMLLYLLIOVLYESYVNIPLFKRFEVVIY-LILELIL 206
 DB 171 -----IFLSARRSMIAFFVIFLYLFLVLE-----KSKKIKGLAFWSYNEFLII 213
 QY 207 ISFLSYT-----KOKVONEIMAKOEQINRITQSOQIESLKDIRSPHDLNLT- 258
 DB 214 LSPFIGGVYLSQKHFKLNDLILGK-----BLNQNLNLSISSGLNLL 260
 QY 259 ----TSLRLGLENKDLASIEKIYHQL--EKTGHQLODTRYNIGHLANIQNDVAKGIL 310
 DB 261 EGISIIKEDIENKRFNL-LIGHGVAGEYMPHRMGTQHRYSIFIVSEFLERGITL 316
 RESULT 15
 YC26_PORPU STANDARD; PRT; 656 AA.
 AC P51392;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical sensor-like histidine kinase ycf26 (EC 2.7.3.-).
 GN ycf26.
 OS Porphyra purpurea.
 OC Chloroplast.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 OX NCBI_TaxID=2787;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AVONPORT;
 RA Reith M.E., Munholland J.;
 RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast genome."
 RL Plant Mol. Biol. Rep. 13:333-335(1995).
 CC -1- SUPRACELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN; CHLOROPLAST (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL, U38804; AAC08278.1; -
 DR InterPro: IPR004358; BCTRLSENSOR.

DR InterPro: IPR003660; HAM.
 DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR InterPro: IPR003661; His_kinA.
 DR InterPro: IPR000014; PAS.
 DR Pfam: PF00672; HAM; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR Pfam: PF00989; PAS; 1.
 DR Pfam: PF00512; signal; 1.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR SMART: SM00304; HAM; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR SMART: SM00388; HISKA; 1.
 DR SMART: SM00091; PAS; 1.
 DR PROSITE: PS50109; HIS_KIN; 1.
 DR PROSITE: PS50112; PAS; 1.
 KW Hypothetical protein; Sensory transduction; Transferase; Kinase;
 KW Transmembrane; Phosphorylation; Chloroplast.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT DOMAIN 289 359 PAS.
 FT MOD_RES 424 654 HISTIDINE KINASE.
 FT MOD_RES 427 427 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 656 AA; 75362 MW; 2B3DDE7C91AE680 CRC64;

Query Match 5.3%; Score 116; DB 1; Length 656;
 Best Local Similarity 20.5%; Pred. No. 5.9;
 Matches 68; Conservative 68; Mismatches 110; Indels 86; Gaps 15;

QY 161 IPMNTMLLYLLIYVL-----VYESYVITFLK--FRK-FVYVYVIL--FLIL- 206
 DB 340 LPEDINQQLPFLINDIKKNFLBEOISICETQICIKLQKNKTKFRVLLTVLDHKYSILK 399
 QY 207 -ISFLSYTKKQVQ-NEIMAKOEQINRITQSOQIESLYKDIRSF-----RHAYLNTL 258
 DB 400 GIAMTQDRTQVEFLNETKNO-----FISVSHELRTPLFNIRSFLYETLYEHYHDLSDS 453
 QY 259 TSLR-LGLENKDLASIEKIYHQLLEKTHQLODTRYNIGHLANIQNDVAKGIL 305
 DB 454 QKLEFLANAKETGRLVNDVLD-----LSRLSPQETTLQPTLVSA 498
 QY 306 VSGLSAKILEAKNKIIVANVEVSSKIQLPPEMLDFTLLSLCDNAIPA----- 356
 DB 499 VQOTIRTYOISAKRIDLHIDIEONLQ---CVLGNVNLLOLIANLVVSLKFTAPNGI 555
 QY 357 -----AFESLNPETIQALAFPKKNGSIYFIIONSTRKEROIVSKIFK-----ENYSTKGSN 405
 DB 556 ILKRAYYDCLKTEYEVGFHNSQKRYVEICDNGIGISRKQDERIFAFRLIENYVHTLGS 615
 QY 406 RIGIGLAKVNHILIEHYPKTSLOTSNHHHLKOL 437
 DB 616 TGIGLSIVAKNIQKH-----NSEIHLYSSEL 640

Search completed: July 30, 2002, 15:20:54
 Job time: 365 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:20:30 ; Search time 45.14 Seconds

(without alignments)
1690.092 Million cell updates/sec

Title: US-09-833-017-4
Sequence: 1 MNEALMILSNGLITVLTFLV.....KTSLSQTSNHHHLFKQLLIK 441

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	773	35.6	446	2	Q9F2F2 streptococ
2	755	34.7	446	2	Q9F2F5 streptococ
3	749	34.5	446	16	Q9S115 streptococ
4	545.5	25.1	426	2	Q32832 streptococ
5	539.5	24.8	448	16	Q9A1J7 streptococ
6	526	24.2	451	16	Q9YPA4 streptococ
7	517	23.8	427	16	Q9ALJ6 streptococ
8	455	20.9	447	16	Q97M11 clostridiu
9	455	20.9	447	2	P95767 streptococ
10	450	20.7	453	2	P72446 streptococ
11	449.5	20.7	453	2	Q33646 streptococ
12	443.5	20.4	448	2	Q33597 streptococ
13	439.5	20.2	441	2	Q9X995 streptococ
14	438.5	20.2	441	2	Q9X990 streptococ
15	437	20.1	452	2	P72444 streptococ
16	436.5	20.1	440	2	Q33616 streptococ

17	436.5	20.1	441	2	Q33652 streptococ
18	435.5	20.0	441	2	Q33667 streptococ
19	434.5	20.0	441	2	Q33669 streptococ
20	432.5	19.9	441	2	Q9X9B8 streptococ
21	427.5	19.7	440	2	Q33671 streptococ
22	427.5	19.7	441	16	Q9S1K4 streptococ
23	426.5	19.6	441	2	Q9X9B6 streptococ
24	426	19.6	441	2	Q33673 streptococ
25	424.5	19.5	441	2	Q9X9B4 streptococ
26	423.5	19.5	441	2	Q9J822 streptococ
27	420.5	19.4	441	2	Q9X9B2 streptococ
28	416	19.1	431	16	Q92FRO listeria in
29	414.5	19.1	441	2	P72566 streptococ
30	411.5	18.9	429	2	Q9S3T7 streptococ
31	394	18.1	425	2	Q33590 streptococ
32	389	17.9	437	2	Q33587 streptococ
33	381	17.5	430	2	Q52182 streptococ
34	379	17.4	423	2	Q33644 streptococ
35	378	17.4	411	2	Q9F622 streptococ
36	378	17.4	429	2	Q68159 streptococ
37	377	17.3	430	2	Q07911 streptococ
38	375	17.3	425	2	Q9F1U4 streptococ
39	375	17.3	425	2	Q9F1U2 streptococ
40	375	17.3	430	2	Q52180 streptococ
41	374	17.2	430	2	Q52181 streptococ
42	372	17.1	371	16	Q9S5L2 streptococ
43	366.5	16.9	427	2	Q9X473 streptococ
44	362	16.7	428	2	Q48866 lactobacil
45	362	16.7	430	2	Q48874 lactobacil

ALIGNMENTS

RESULT 1
ID Q9F2F2 PRELIMINARY; PRT; 446 AA.
AC Q9F2F2;
DT 01-MAR-2001 (TREMUREL. 16, Created)
DT 01-MAR-2001 (TREMUREL. 16, Last sequence update)
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)
DE HISTIDINE KINASE.
GN HK.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20487097; PubMed=11034284;
RA Reichmann P., Hakenbeck R.;
RT "Allelic variation in a peptide-inducible two-component system of Streptococcus pneumoniae."
RL FEMS Microbiol. Lett. 190:231-236(2000).
DR EMBL: AJ278302; CAC10175.1; -
DR InterPro: IPR001064; CrystalLin.
DR InterPro: IPR003594; HATPase_C.
DR SMART: SM00387; HATPase_C.1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN.1.
KW Kinase.
SQ SEQUENCE 446 AA; 52057 MW; 98065278BB9C2C1E CRC64;

Query Match 35.64; Score 773; DB 2; Length 446;
Best Local Similarity 38.84; Pred. No. 5.4e-37;
Matches 174; Conservative 91; Mismatches 168; Indels 16; Gaps 6;

QY 1 MNEALMILSNGLITVLTFLVLFPSKVSNTSKRELFLFSISNPLIMAVT-----M 54
DB 1 MNVAMIL-----LYLVTHGKLIVIFKVDGISLTFER--IFKALFLKILLAVFGMLGYM 54
QY 55 V-NVNLFPAPPLVFIALSTVILNRQNSLSLNFYGLLPVASSDLFPRAILFIID--GTQ 111

```

Db 55 VGNVYLSFEMPRLVIGTIGSLLELRLKLLFGLFPLLVNLFVRGVSFFLPLRGG 114
Oy 112 GIVMGSSITTYTMEFAGIALSYLFSVFNVDIGRLKDSLTJKKVKRLIPMNTMLLY 171
Db 115 QYVDYSFIMWCIITF--NFIISLAFKMLVDYDFSLFKGILDKDPKSLQJWIMGAVY 173
Oy 172 LLIQVLYIESYVNPITLKFPEYVIYLLIFLLISLSQYTKQVQNMOKRQAOJR 231
Db 174 LVIQMLSTFEYEGQISTYVHLLVLYLFLFMGIKLLTYLKDKLHENOQDLRR 233
Oy 232 NITQYSOQIESLYKDIRFRHDYNIITSLRLGIENKDLASIEKTYHQILEKTHOLODT 291
Db 234 EMERSRHIEELYEKVSFRHDYTNLLTSLRLGIEEDMQIKREIYDVLKOSSEKLODN 293
Oy 292 RYNGHNLANTQNDVAKGLSLAKTIEAQNKKIYAVNEVSSKIQLPMEILLDTITLSILD 351
Db 294 KYDQRLVNDRAKSLSLACKFTKANDKNIYVNEVPEIQYEGVSLDELTVVSLD 353
Oy 352 NAIEAFESLNPETQLAFKNGSIVFIQNSTKEKQIDVSKIFKENYSTKSGNRGIGLA 411
Db 354 NAIEASVACOPHVSIAFFKNGAOETFIENSIREGIDISEIFSGASSKGEERGVLTY 413
Oy 412 KVNHLHYPKTSLOTSNNHLLFKOLLII 440
Db 414 TVMKIVESHPTSLNTTCQNVFROVLTV 442

RESULT 2
Oy 25 PRELIMINARY: PRT: 446 AA.
ID 09F2F5
AC 09F2F5:
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE HISTIDINE KINASE.
GN IH.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-R6;
RX MEDLINE=20487097; PubMed=11034284;
RA Reichmann P., Hakenbeck R.;
RT "Allelic variation in a peptidase-inducible two-component system of
RT Streptococcus pneumoniae."
RL FEMS Microbiol. Lett. 190:231-236(2000).
DR InterPro; IPR001064; CysTallin.
DR SMART; SM00387; HATPase_C.1.
DR PROSITE; PS00225; CRYSTALIN_BETAGAMMA; UNKNOWN_1.
KW Kinase.
SQ SEQUENCE 446 AA; 51670 MW; C13A9CD1EAEE22C CRC64;

Query Match 34.7%; Score 755; DB 2; Length 446;
Best Local Similarity 39.1%; Pred. No. 5.8e-36;
Matches 175; Conservative 88; Mismatches 171; Indels 14; Gaps 5;

```

```

Oy 173 LLIQVLYIESYVNPITLKFPEYVIYLLIFLLISLSQYTKQVQNMOKRQAOJR 232
Db 175 VMQSLSTFEYEGQISTYVHLLVLYLFLFMGIKLLTYLKDKLHENOQDLRRD 234
Oy 233 ITQYSOQIESLYKDIRFRHDYNIITSLRLGIENKDLASIEKTYHQILEKTHOLODT 292
Db 235 EMERSRHIEELYEKVSFRHDYTNLLTSLRLGIEEDMQIKREIYDVLKOSSEKLODN 294
Oy 293 YNIGHNLANTQNDVAKGLSLAKTIEAQNKKIYAVNEVSSKIQLPMEILLDTITLSILD 352
Db 295 YDLQRLVNDRAKSLSLACKFTKANDKNIYVNEVPEIQYEGVSLDELTVVSLD 354
Oy 353 AIEAFESLNPETQLAFKNGSIVFIQNSTKEKQIDVSKIFKENYSTKSGNRGIGLA 412
Db 355 AIEASVACOPHVSIAFFKNGAOETFIENSIREGIDISEIFSGASSKGEERGVLTY 414
Oy 413 VNHLEHYPKTSLOTSNNHLLFKOLLII 440
Db 415 VMKIVESHPTSLNTTCQNVFROVLTV 442

RESULT 3
Oy 315 PRELIMINARY: PRT: 446 AA.
ID 09S115
AC 09S115:
DT 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DT 01-OCT-2001 (Tremblrel, 18, Last annotation update)
DE HISTIDINE KINASE (BLPH PROTEIN) (SENSOR HISTIDINE KINASE BLPH,
DE PUTATIVE).
GN HK13 OR IH OR BLPH OR SP0527.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99453779; PubMed=10524254;
RA Lange R., Wagner C., de Saizieu A.B., Flint N., Molnos J., Stieger M.,
RA Caspers P., Kamber M., Keck W., Amrein K.;
RT "Domain organization and molecular characterization of 13 two-
RT component systems identified by genome sequencing of S. pneumoniae."
RL Gene 237:223-234(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-KNR7/87;
RA Reichmann P., Hakenbeck R.;
RT "A peptidase inducible Signal Transduction System in Streptococcus
RT pneumoniae: Evidence for Bacteriocin Production."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBD databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Debey R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).

```

DR EMBL: AJ006401; CAB54587.1; ...
 DR EMBL: AJ278419; CAC18581.1; ...
 DR EMBL: AJ278410; CAC03516.1; ...
 DR EMBL: AE007364; AAK74685.1; ...
 DR TIGR: SP0527; ...
 DR InterPro: IPR001064; Crystalin.
 DR PROSITE: PS00225; CRYSTALLIN_BETA; UNKNOMN_1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 446 AA; 51824 MW; D8CB0D28P5D274C CRC64;

Query Match 34.5%; Score 749; DB 16; Length 446;
 Best Local Similarity 37.8%; Pred. No. 1.3e-35;
 Matches 170; Conservative 89; Mismatches 173; Indels 18; Gaps 4;

QY 1 MNEAMML-----SNGLLTYLVLEFLFESKSNVTL-----KKEFLFESINFLMIA 51
 DB 1 MNTAMLLYLVNGL-----ELVIFKVDGIDLFERIFKELLKLLAEVAMIS 52
 QY 52 VTWVNVLPYAPLPYFALSTYLNKNSLSNIFVGLPVASDLPFRRAIFFILDGT- 110
 DB 53 YVGNVYLSYFMEPLVIGLSFLRLGLPKLLFEGLEPMLVNLFYRGVSFVLPGLG 112
 QY 111 OGIVSSITTYMIEFAGIALSYFLSVFNDIGRLDSLRKMKVKKRLIPNITMLY 170
 DB 113 OGQVYDGYSTGLCTIIFNFFSLAFKWLDDFTSLKREIDKRFQSLQIMMIGY 172
 QY 171 YLLQLVLYIESYVNIPTLKFKEFVYVYLLLELISFLSQYTRKQVONEIMAKENQI 230
 DB 173 YLVWESLSEFVEYQSIQSKVRLHLLVFFLLPFMGVYKLDLYLEKRYEREDQALRY 232
 QY 231 RNIVYQSOEESLYKDIRSPRHDYNTLTLRGTENKDLASIEKIYQILEKQHOD 290
 DB 233 RDMRYRRIHELEKVEKRSRHDYNTLTLRGTENKDLASIEKIYQILEKQHOD 292
 QY 291 TRNIGHLANTONDAVKGISAKILEONKKAIVNVESSKIQLEPEMLDPTILSLC 350
 DB 293 NKDGLVLRKALKSLAGKFLKAKDNIVFENVEEIOVGMISLDFLIVSLIC 352
 QY 351 DNAIEAFESLNEIOLAFKFKNSIVFTIONSTKEKOJDIKREKYNSTGSKNRIGL 410
 DB 353 DNALASVEASOPHVSJAFKNGAQETFLIENSKEGIDISEIFGASNGEERGVL 412
 QY 411 AKVNHLEHFKPSLOTSNHHHLEFKOLL 440
 DB 413 YTWKIVESHPTSLNTTCOMQVRFQVLT 442

RESULT 4
 ID 032832 PRELIMINARY; PRT: 426 AA.

AC 032832; 01-JUN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PUTATIVE HISTIDINE KINASE (EC 2.7.3.-) (ORF8).
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 OC Lactobacillus
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LPC010;
 RX MEDLINE=98247325; PubMed=9572965;
 RA Stephens S.K., Floriano B., Cathcart D.P., Bayley S.A., Witt V.F.,
 Jimenez-Diaz R., Warner P.J., Ruiz-Barba J.;
 RT "Molecular analysis of the locus responsible for production of
 plantarum LPC010", a two-peptide bacteriocin produced by Lactobacillus
 plantarum LPC010".
 RL Appl. Environ. Microbiol. 64:1871-1871(1998).
 CC -1- SUBCELLULAR LOCATION: INTERAL MEMBRANE PROTEIN (POTENTIAL).
 DR EMBL: Y15127; CAA5398.1; -

DR InterPro: IPR003594; HATPase_C.
 DR InterPro: IPR004359; HIS_KIN_sig.
 DR Pfam: PF02518; HATPase_C; 1.
 KW Kinase; Transferase; Transmembrane.
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 78 98 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 SQ SEQUENCE 426 AA; 48387 MW; 4B203DF09EC34C0 CRC64;

Query Match 25.1%; Score 545.5; DB 2; Length 426;
 Best Local Similarity 30.3%; Pred. No. 5.7e-74;
 Matches 131; Conservative 99; Mismatches 174; Indels 29; Gaps 7;

QY 18 VLFLPLFSKSNVTLSEKELTESINFLMIAVNVNLEPPAPLPYFALSTYLNK 77
 DB 11 VTSLLIYHRSISIKLTMREIALVLELGFSSILS---TFEFAVIMMLTAIGIKDGL 66
 QY 78 QNSLNIYFGLPVASDLPFRRAIFFILDGTQGIYMS---SIITTYMIEFAGIAL 132
 DB 67 HQKYLAEFFSYV-----SVTEFSLVGNLSLGFVGDENIYIPFTALPVI 116
 QY 133 SYLFSEVNVNIGRLKDSLRKMKVKKRLIPNITMLYLLVLIQVLYISYVNIPTL 189
 DB 117 NELICKLNPDPDFLRNV--GATRNLYGVN---FIFSLCTVQY--GSYWEEDMGDI 170
 QY 190 -KRRFVYVYLLLELISFLSQYTRKQVONEIMAKENQINIVYQSOEESLYKDIR 248
 DB 171 GYVRVYLGSLFLLIALVLIYLNKLSAQOQSLRCREGRQYHMQYSQLEINLTVEVR 230
 QY 249 SPRHDYNTLTLRGTENKDLASIEKIYQILEKQHODRNVNIGHLANTONDAVKG 308
 DB 231 SRHDYNTLTLRGTENKDLASIEKIYQILEKQHODRNVNIGHLANTONDAVKG 290
 QY 309 IISAKILEONKKAIVNVESSKIQLEPEMLDPTILSLCNDALBAFESINPEIOLA 368
 DB 291 ILSKTIYAQONGIEVNISSDVTTYFDLDIRVIAFLNALIASLTDLPMSVA 350
 QY 369 FPKKNSIVFTIONSTKEKOJDIKREKYNSTGSKNRIGLAKVNHLEHFKPSLOTS 428
 DB 351 FIEDDSQVLIYVNNAPSYIDKRRIFESGSTGNHKGIGLATVMDLNKRYNASTLEFE 410
 QY 429 NHHHLEFKOLL 441
 DB 411 FRNSLEFQKLAIR 423

RESULT 5
 ID 09A1J7 PRELIMINARY; PRT: 448 AA.

AC 09A1J7; 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PUTATIVE HISTIDINE KINASE, POSSIBLY INVOLVED IN COMPETENCE.
 GN SPY0242.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE=21192684; PubMed=11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
 Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes".
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 CC EMBL: AE006491; AAK33320.1; -
 DR InterPro: IPR003594; HATPase_C.


```

QY 63 ---AEPYIATSLYNQNSLSNIFGCLPVASSDLFRRAIFPILDTGTGWSST 119
DB 52 VILIDLEFIVLSAPKKLEFMNGEFTLIVLELVISFPLPVLGSIQIN 111
QY 120 ITYVMEFA---GIALSYLSEVFNVDIGRLKSLTKKAKKRLIPNNITMLLYLLIQV 176
DB 112 NNKLELCYLVLPFVFLFYFISIDLSLR-FISECKMKKVMFNMTAFSYFAHF 170
QY 177 LVIESYNTPLTKRREYVYLLFLILISLQTKQVQNEIMAKOEQIRNITQY 226
DB 171 LVTVSGFALFYQYRSLVFLYLAIFTWIVKIDRFKQDLSQLTQAOERLAVLENY 230
QY 237 SQOIEYLDIRSFHDYLNITLSRLGIENKDLASIEKIYHQLIEKTHOLOPTRNIG 296
DB 231 NQSIQLEYREIVKVDHSENLISLKDSDIDDLITRVYDVYIQQSATSMRTNYS 290
QY 297 HIANONDAVKILSAKILEAQNKIAVNEVSSKIQLEPEMLDFTIISLCDNAIEA 356
DB 291 SLDNIEAVIRSMNSKLEAOYIGIEYIEIPVDIHLPLKIDLVLEFTGLVDNIET 350
QY 357 AFESLNPETQLAFKNGSIVFIIONSTRKQIDVSKIFENYSTKSGNIGLAKVNH1 416
DB 351 AKGSRPEFSIAYFKQDMKQLEFIENSTKTRNVDIAKFAQOQNS-----AHLIV 402
QY 417 LEHYPTSLQTSNHHLEKOLL 438
DB 403 LQSPQITVSTKSDHRLKOLL 424

RESULT 8
Q97MM1 PRELIMINARY: PRT: 447 AA.
AC 097MM1:
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HISTIDINE KINASE-LIKE AMPEASE;
GN CAC0080.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VYM B-1787;
RC MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Olu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007520; AK78065.1;
DR InterPro: IPR003594; HATPase_C;
DR SMART: SM00387; HATPase_C;
DR KINASE: Complete proteome.
SQ SEQUENCE 447 AA; 51289 MM; A0352A687A692F7B CRC64;

```

```

Query Match 20.9%; Score 455; DB 16; Length 447;
Best Local Similarity 27.4%; Pred. No. 9.4e-19;
Matches 124; Conservative 96; Mismatches 163; Indels 70; Gaps 10;

QY 11 GLTYLVFLFLFESKVSNTLSKELTFSISNFLIMAVTNVNLFPAPAEFIA 70
DB 33 GMVITEIISLFLM-----LGMQYVMA-----LII 61
QY 71 LSTYLRKNS---LSLNTYGLPVASS-----DLFRAIIFPILDTGTGWSST- 120
DB 62 PCIFMTLQTSDDVLSIS---LPVAGCLIIIVMDYFLSVYVNVFVSVIKNDSYIH 116
QY 121 -TYVMIEFAGIALSYLSEVFNVDIGRLKSLTKKAKKRLIPNNITMLLYLLIQVLYV 179

```

```

DB 117 WITFCIEFTGYVSTRIRKFEVD-----KFRAYNGHKGITGVAAAGTLLIMFV 167
QY 180 IESYVNI-----PTLKRFRKVYV-YLLIFLILISLSQYTKQVQNEIMAKOE 228
DB 168 FYNNVAVPAPNSVNSNTWRKIGVLFESYAILLIIVIRTLIRGIRKME---LKS 224
QY 229 QIRNITQYSQOIESLYNDISFRDYINILTSRLGIENKDLASIEKIYHQLIEKTH 288
DB 225 EFQSLQETYNLEKTLHKDMGFPRDYINILSMAGYIONRDEGLEFRFDDKIMPLSKAM 284
QY 289 QDTRYNIGHLANIONDAVKILSAKILEAQNKIAVNEVSSKIQLEPEMLDFTIISL 348
DB 285 KSNNEKIDLLQINIEVPEIKGFWSAKIRAQETGIDYIDVAESIKSFNEIIDL 344
QY 349 LCDNAIEAFESLNPETQLAFKNGSIVFIIONSTRKQIDVSKIFENYSTKSGNIGI 408
DB 345 LLDNAIEASEKCDRPSMKVAIVNKKDSVMVILNNNEIPIYKIKRGSTGDKDNKI 404
QY 409 GLAKVNHILEHYPTSLQTSNHHLEKOLLIK 441
DB 405 GLSNLKDIIGKYPVMDLVTEIENQFQKIIDIK 437

RESULT 9
P95767 PRELIMINARY: PRT: 453 AA.
AC P95767:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HISTIDINE PROTEIN KINASE.
GN COMD.
OS Streptococcus gordonii challis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=29390;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CHALLIS;
RC MEDLINE=97294448; PubMed=9150204;
RA Lunford R.D., Robie A.G.;
RT "comY, a gene similar to comGA of Bacillus subtilis, is essential for
RT competence-factor-dependent DNA transformation in Streptococcus
RT gordonii.";
RL J. Bacteriol. 179:3122-3126(1997).
DR EMBL: U80077; AAC45306.1;
DR KINASE.
SQ SEQUENCE 453 AA; 53636 MM; 781D5360489B0139 CRC64;

```

```

Query Match 20.9%; Score 455; DB 2; Length 453;
Best Local Similarity 28.9%; Pred. No. 9.5e-19;
Matches 131; Conservative 94; Mismatches 179; Indels 50; Gaps 12;

QY 1 MNEALMISNGLTYLVFLFESKVSNTLSKELTFSISNFLIMAVTNVNL 59
DB 32 INIKIILFKWILWISFLYSLF-----LLPYMKTRIFRIFRYLL----- 74
QY 60 FYPAEPLFYALSILYNRONSLSNIFGL---LPVASSDLFRRAIIFPILDTGTGIW 116
DB 75 ---IPLFYRYFFPKRREYVYLLFLSEVYLSVETSETFLSVIT-----SSITG 123
QY 117 SSIYTT-YMIEFAGIALSYLSEVFNVDIGRLKSLTKKAV-KRRLIPNNITMLLYLLI 174
DB 124 DSFVQOHDIRIYIILNLSLFLILKVVDPDFEYFYEYKKEIYKNDLYNNKSYIVIHLL 183
QY 175 QVLY-----VIESY-NVITPLKRFVYVYLLFLILISLQTKQVQNEIMAKOE 227
DB 184 NISHFESSENAHNSFASATATIGF-----IMFSLTFYKASAREYKAK---ETQ 235
QY 228 AQIRNITQYSQOIESLYNDISFRHDYLNITLSRLGIENKDLASIEKIYHQLIEKTH 287

```

Db 236 EQ-ROLQYLTDEIVGLVNEIRGFRHDYAGMLTSLQTSINGSDMKVEYRIFHNVLQSANIS 294

QY 288 LQDTRYNIGHLANIQNDVAKGILSAKLLEAQNKRIAVNEVSSKIOLPEMELLDTITIS 347

Db 295 LRSDDYTFEFLNNVODTALRSVLQITIFKARECGVEIYFEKKDVIEFLPKMLDLVAVAS 354

QY 348 ILCDNAIEAFESLNPETQLAFPKNGSIYFIIONSTREKOIDVSKIFKENSSTKGSNG 407

Db 355 VILNNVAGAAESPSTKMNVLVLDEKEIVFIIONSRQSRITNLEIYEVGFSTKGENRG 414

QY 408 IGLAKVNHILEHYPKTSLOTSNHHHLFKOLLITK 441

Db 415 IGLNNVKEIIDKYDEVILLETDIETNYFIQVVRFK 448

RESULT 10

P72446 PRELIMINARY; PRT; 453 AA.

AC P72446 01-FEB-1997 (Tremblrel. 02, Created)

DC 01-FEB-1997 (Tremblrel. 02, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE HISTIDINE KINASE.

GN COMD.

OS Streptococcus gordonii.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1302;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CHALLIS (NCTC 7868);

RX MEDLINE=97032151; PubMed=8878047;

RA Haverstein L.S., Gaustad P., Nes I.F., Morrison D.A.;

RT "Identification of the streptococcal competence-pheromone receptor.";

RL Mol. Microbiol. 21:863-869(1996).

DR EMBL; X98109; CAA66787.1; -.

KW kinase.

SQ SEQUENCE 453 AA; 53633 MW; 281D55CDE4063E2 CRC64;

Query Match 20.7%; Score 450; DB 2; Length 453;

Best Local Similarity 28.6%; Pred. No. 1.8e-18;

Matches 130; Conservative 94; Mismatches 180; Indels 50; Gaps 12;

QY 1 MNEALMILSNGLLTYLVFLFESKYSNVLSSKELTLF-SISNPLIMATMVNVL 59

Db 32 INIKIILFWMGLIMVLSFLSLF-----LIPDYMKIFRDIHFHFL----- 74

QY 60 FVPAPFLYFIALSYLNRONSLSNIFYGL--LPVASSDLFRAIIFELIDGTGIVMG 116

Db 75 ----IQPLFYKFLIKRREYENYVNLFLFYIYLSVETSTFSLVIL-----SSITG 123

QY 117 SSITTT-WIEPAGIALSYLFSVFNVDIGRLKDSLTKMKV-KRRLPMNITMLLYLLT 174

Db 124 DSVFKOHYDIFLIINLFLFIILKVVDFEFYEFYKPEYKNDLVNMSYVLIHL 183

QY 175 OVLY-----VIESY-NVLPFLKFRFVYIYLLFLILISLSQYTKQKQVEMAKQE 227

Db 184 NISHMSENAHLNFSAMATIGF-----IMFSTLFLKSAREQYERAK---EIQQKE 235

QY 228 AQIRNITQYSOQIESLYKDIRFHDYNIILTSRLGTEKNDLASIEKTYHOLEKTHQ 267

Db 236 EQ-ROLQYLTDEIVGLVNEIRGFRHDYAGMLTSLQTSINGSDMKVEYRIFHNVLQSANIS 294

QY 288 LQDTRYNIGHLANIQNDVAKGILSAKLLEAQNKRIAVNEVSSKIOLPEMELLDTITIS 347

Db 295 LRSDDYTFEFLNNVODTALRSVLQITIFKARECGVEIYFEKKDVIEFLPKMLDLVAVAS 354

QY 348 ILCDNAIEAFESLNPETQLAFPKNGSIYFIIONSTREKOIDVSKIFKENSSTKGSNG 407

Db 355 VILNNVAGAAESPSTKMNVLVLDEKEIVFIIONSRQSRITNLEIYEVGFSTKGENRG 414

QY 408 IGLAKVNHILEHYPKTSLOTSNHHHLFKOLLITK 441

Db 415 IGLNNVKEIIDKYDEVILLETDIETNYFIQVVRFK 448

RESULT 11

O33646 PRELIMINARY; PRT; 453 AA.

AC O33646 01-JAN-1998 (Tremblrel. 05, Created)

DC 01-JAN-1998 (Tremblrel. 05, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE HISTIDINE KINASE.

GN COMD.

OS Streptococcus gordonii.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1302;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 3165;

RX MEDLINE=98012953; PubMed=9352904;

RA Haverstein L.S., Hakenbeck R., Gaustad P.;

RT "Natural competence in the genus Streptococcus: Evidence that streptococci can change phenotype by interspecies recombinational exchanges.";

RL J. Bacteriol. 179:6589-6594(1997).

DR EMBL; AJ000870; CAA04357.1; -.

KW kinase.

SQ SEQUENCE 453 AA; 52756 MW; ABF273293A2CC52 CRC64;

Query Match 20.7%; Score 449.5; DB 2; Length 453;

Best Local Similarity 28.2%; Pred. No. 2e-18;

Matches 127; Conservative 91; Mismatches 173; Indels 59; Gaps 10;

QY 12 LITVITVFLFLFESKYSNVLSSKELTSLSNPL--IMATMVNVLFPAPFLYRI 69

Db 38 IFVLTTFESSLITQFYVNLPEYGDV---LSFPLHYLILFQPLILHFSKGLKYG 93

QY 70 ALSTYLNRRNSLSNIFYGLLPVASSDLFRAIIFELIDGTGIVMGSSITTY----- 123

Db 94 YVSLP-----LSLIY---LSVSSSTFSSVIL-----SSVNGDFVNOVMSYYT 136

QY 124 -----MIEFAGIALSYLFSVFNVDIGRLKDSLTKMKVKKRRLIPMNITMLLY 171

Db 137 IVNIALFVFLKSEFEFEFSYKNDFEKEI-----MNVIKIYTIITLINISH 187

QY 172 LLIQVLYIESYNYVLPFLKFRFVYIYLLFLILISLSQYTKQKQVEMAKQAIR 231

Db 188 WFSNAHLNFSAMATIGF-----IMFSLTFYKLSAREQYERAK---EIQQKEEQ-R 238

QY 232 NITQYSOQIESLYKDIRFHDYNIILTSRLGTEKNDLASIEKTYHOLEKTHQLODT 291

Db 236 EQ-ROLQYLTDEIVGLVNEIRGFRHDYAGMLTSLQTSINGSDMKVEYRIFHNVLQSANIS 294

QY 292 RYNIGHLANIQNDVAKGILSAKLLEAQNKRIAVNEVSSKIOLPEMELLDTITISLCD 351

Db 299 DTFEFLNNVODTALRSVLQITIFKARECGVEIYFEKKDVIEFLPKMLDLVAVASVLLN 358

QY 352 NATEAFESLNPETQLAFPKNGSIYFIIONSTREKOIDVSKIFKENSSTKGSNGIGLA 411

Db 359 NAVEGAASPSKTMNVLVLDEKEIVFIIONSRQSRITNLEIYEVGFSTKGENRGGLIN 418

QY 412 KVNHILEHYPKTSLOTSNHHHLFKOLLITK 441

Db 419 NKETIIDKYDEVILLETDIETNYFIQVVRFK 448

RESULT 12

O33597 PRELIMINARY; PRT; 448 AA.

AC O33597 01-JAN-1998 (Tremblrel. 05, Created)

DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE HISTIDINE KINASE.
 GN COMD.
 OS Streptococcus anginosus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 RX NCBI_TaxID=1328;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NTC 10713;
 RX MEDLINE=98012953; PubMed=9352904;
 RA Havarstein L.S., Hakenbeck R., Gaustad P.;
 RT "Natural competence in the genus Streptococcus: Evidence that
 streptococci can change phenotype by interspecies recombinational
 exchanges.";
 RL J. Bacteriol. 179:6589-6594(1997).
 DR EMBL; AJ000864; CNA0342.1;
 DR InterPro; IPR003594; HATPase_C.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR Pfam; PF02518; HATPase_C; 1.
 KM Kinase.
 SQ SEQUENCE 448 AA; 52922 MW; A995DB9B254ACB57 CRC64;

 Query Match 20.4%; Score 443.5; DB 2; Length 448;
 Best Local Similarity 23.6%; Pred. No. 4.3e-18;
 Matches 105; Conservative 113; Mismatches 195; Indels 31; Gaps 8;

 QY 12 LLYLVVLFLEFSKVVSVNTLSKKEITFSTSNFLIMATVMNVNLFYPAE-----PL---YFALSTYLNKONS 80
 DB 10 IFETIIRAVSVYCYKKNVKNVNIHFTLCIGIVFSTDFTLHYSTRVMFFIQLFF 69
 QY 69 IALSTYLNKONSLSNIFGL--LPVASSDLFRRAIFPILDTGQIVMGSSITITTYMI 125
 DB 70 Y-YFFKVKVKKKHLVFLALFLSLAVSSSEFFSVIT-----SSVTGDKFVDRW- 119
 QY 126 EFAGIALSYLFLSVN-----VDIGRLK-DLSTKMKVKKRLIPNITMLLYLLIOVL 177
 DB 120 -----GLFYFNIISLVFLKAIIDFKNFYFKKIDFKKEIYQALNLYLLHLLNVS 174
 QY 178 YVIESVNAVITLTKRFKVVIVYLLIFLLISLSQTYKQVONE-MAKQEQINITYS 237
 DB 175 HWLS--NMKNLNSFSSMNTICFLMMSILFLQSIKREYKEEQIKKEEQQLQKYT 232
 QY 238 QQIESLYKDIRSPRDYLLITSLRIGIENKDLASIEKIYHQLIEKTHQLODTRVNI 297
 DB 233 DELVSYLNEIRGRHDYGMASFOGAIHTGDIKEVERIYGVVLANMLQLRSDKYTFD 292
 QY 298 LANIQNDVAVGILSAKILEKQNKKTIVNVEVSKIQLPFEMELDTITLSLCDNAIEA 357
 DB 293 LNNVGSALMSVMTQYLFTRDNIETLFEVDFVNPFLKILDLVRRMYSVLLNNALEGA 352
 QY 358 FESLNPETIOLAFPKKNGSIYFIIONSTREKQIDVSKIFENYSTKSGNIGIGLAKVNHIL 417
 DB 353 AESYQKTMNYSVLVDLTETELIVIONSKRRPRDLEIYQTDSTSGEGRGIGLSNIKEII 412
 QY 418 EHYKPTSIQPSNNHHLFKOLLIK 441
 DB 413 NNYGIIIDTKIEDFTFOVMRYR 436

 RESULT 13
 ID Q9X995 PRELIMINARY; PRT; 441 AA.
 AC Q9X995;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE COMD PROTEIN.
 GN COMD.
 OS Streptococcus oralis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 RX NCBI_TaxID=28037;
 RN [1]

OC Streptococcus.
 RX NCBI_TaxID=1303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COL19;
 RX MEDLINE=99255539; PubMed=10322016;
 RA "Genetic diversity of the streptococcal competence (com) locus.";
 RT J. Bacteriol. 181:3144-3154(1999).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COL19;
 RX MEDLINE=20098376; PubMed=10632854;
 RA Witzell H.;
 RT "Chloroplast DNA variation and reticulate evolution in sexual and
 apomorphic sections of dandelions.";
 RL Mol. Ecol. 8:2023-2035(1999).
 DR EMBL; AJ240794; CAB99420.1;
 DR InterPro; IPR004359; HIS_KIN_sig.
 SQ SEQUENCE 441 AA; 51328 MW; 63AC032CMA3DEC9D CRC64;

 Query Match 20.2%; Score 439.5; DB 2; Length 441;
 Best Local Similarity 27.6%; Pred. No. 7.1e-18;
 Matches 121; Conservative 89; Mismatches 171; Indels 57; Gaps 11;

 QY 33 LSKKELTFSTSNFLIMATVMNVNLFYPAE-----PL---YFALSTYLNKONS 80
 DB 28 LSKKELTFEGIYTLIVE-AVLELSFYLLYDEIVIERFLFPLGLYSFRIRKQY-ERDRG 85
 QY 81 LSLNIFEGILPVASSDLFRRAIFPILDTGQIVMGSSITITITTYMI 127
 DB 86 LFLSLLSLIVESTIN-----FLSYTFSSITGDNFVSQYHDPFFVYVLYTFVVL 136
 QY 128 AGIALSYLFLSVNVDIGRLKDSLTKKVKKRLIPNITMLLYLLIOVLYESVNAV 187
 DB 137 KIHVYFHELSYFDKDY-----LYPFLKVFALLLHLYVFS-DMVS 179
 QY 188 TLK-FRKFVIVYLLIF--LLISFLSQTYKQVONE-MAKQEQINITYSQQIESL 243
 DB 180 TIKHLNLSFGSILSLVFLSLTLTFANSHKEQEKERIALKOKPFQKHLQNTDDEL 239
 QY 244 YDIQSPRDYLLITSLRIGIENKDLASIEKIYHQLIEKTHQLODTRVNI 303
 DB 240 YNEIRGFRHDYAGMLVSMQMAIDSGDLOEDRIDRYNEVLVANKRLSRSDKYTFD 299
 QY 304 DAVKGIISAKILEKQNKKTIVNVEVSKIQLPFEMELDTITLSLCDNAIEA 363
 DB 300 SALRSLVQSIYVARNNGVEFTLEVKDITKLPFLIDLVIRMSVLLNNAVEGSDSYK 359
 QY 364 ETQLAFPKKNGSIYFIIONSTREKQIDVSKIFENYSTKSGNIGIGLAKVNHILEYPT 423
 DB 360 QMEVAVIKMETETVIVIONSKRMTPSGDLFALGFSTGRNGVGLNNKELLDYNNI 419
 QY 424 SLQTSNNHHLFKOLLIK 441
 DB 420 ILETMEGSTFQILIRK 437

 RESULT 14
 ID Q9X990 PRELIMINARY; PRT; 441 AA.
 AC Q9X990;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE COMD PROTEIN.
 GN COMD.
 OS Streptococcus mitis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 RX NCBI_TaxID=28037;
 RN [1]

